

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

Hidden talent

Hawaiian crow joins
New Caledonian crow
as master tool user **PAGE 403**

BUILT ENVIRONMENT

THE GREAT INDOORS

The science of making homes
and offices more healthy

PAGE 294

AGRICULTURE

GROWING PLACES

Fifty years of data point
to shift in food R&D

PAGE 301

INSIGHT

PROTEIN WORLD

Reviewing the hottest topics
in protein science

PAGE 319

NATURE.COM/NATURE

15 September 2016 £10

Vol. 537, No. 7620



THIS WEEK

EDITORIALS

DRUGS Research down the drain can help to fight crime **p.280**

WORLD VIEW Lifting of US chimaera ban offers opportunity **p.281**



HAWAII Warmer islands spread disease to kill birds **p.283**

Unit of contention

The United States' refusal to use SI units for radiation measurement is confusing and dangerous. It's time to catch up with the rest of the world.

There are two types of nation: those that use the metric system and those that have put a man on the Moon. The reliance of the United States on feet and pounds, along with its refusal to embrace metres and kilograms, baffles outsiders as much as it warms the hearts of some American patriots. But it is time for the country to give up on the curie, the roentgen, the rad and the rem.

Instead, US regulators and scientists should adopt the appropriate SI units for the measurement of radioactivity. They should do so not only for the sake of international harmony, but also to protect the health and safety of US citizens.

After years of wrangling, on 29 September the National Academies of Sciences, Engineering, and Medicine will hold a workshop to discuss whether the United States should adopt the international system of units for radiological measurements. The negotiations will affect everyone from NASA astronauts and air crews to emergency responders.

The rest of the world signed up some time ago. In the 1970s, the International Committee for Weights and Measures adopted a clear set of SI units to describe radiation exposure. The curie, an inspiringly named but clunky measure of radioactivity, was replaced with the becquerel. The roentgen, describing air ionization, became a measurement in coulombs per kilogram. The rad, which quantifies absorbed dose, was superseded by the gray. And the rem, which describes the dose that causes the same amount of biological damage as a rad, was replaced by the sievert.

In case of a nuclear accident, this last quantity is the most crucial. Sieverts capture how people's immediate radiation exposure might translate to future health effects. In 2011, after a tsunami swamped the Fukushima Daiichi nuclear power plant in Japan, the International Atomic Energy Agency and Japanese authorities used sieverts to describe releases of radiation from the three failed reactors.

As fear spread and the public and media clamoured for information, the last thing anybody needed was a load of complicated conversions. It was hard enough for most to sort out the difference between millisieverts and microsieverts, never mind then having to convert those to rems. Yet US officials insisted on generating hazard maps using rems. And that meant that people, including those in the danger zone, could not tell at a glance what was really happening.

Yes, it is possible to use both sets of measures, and to follow the rem numbers with the sievert numbers in brackets. In practice, this is what many US regulatory agencies do. But it is simply too awkward. The Australian government has publicly criticized the US system for creating confusion.

In the middle of an international nuclear-radiation incident, should emergency-response officials huddled in a situation room really need to whip out their calculators? Remember NASA's Mars Climate Orbiter, which was lost in 1999 when someone forgot to convert between imperial and metric units (even though they had plenty of time to check) — the spacecraft broke apart in the Martian atmosphere

rather than smoothly entering orbit. Imagine if such an embarrassing error involved the life and safety of millions of people here on Earth.

Many US experts know that they need to make the switch. Officially, the government encourages agencies to use SI units. And unlike with everyday measures of distance and mass, Americans don't have a deep and lasting emotional bond with radiological measures, and could easily be brought to understand sieverts. During Fukushima, many

US news agencies gave up on even trying to convert, and simply used the international sievert measures.

"In the middle of a radiation incident, should emergency-response officials need to whip out their calculators?"

So why not make the change? The US nuclear industry claims it will be expensive, with millions of dollars needed to update software and hardware and to retrain workers. (In 2012, the country's Nuclear Regulatory Commission, which technically oversees the industry but is widely sympathetic to it, quashed an

effort to switch to SI units.) But the US nuclear industry's suppliers also sell to European manufacturers, and so are well equipped to adapt.

In the eighteenth century, French scientists proposed the metric system, and then French officials imposed it. US researchers should follow their lead, and then US regulators should make the change, and require the industry to follow.

In 1914, an article in *Nature* bemoaned the fact that the metric system was slow in catching on: "Why do people go on agitating? Well, the reason is the necessity for such a system." A century on, the United States is running out of reasons not to bring its radiation measurement into the modern era. ■

No way out

Questions abound over the deportation and subsequent house arrest of a physicist.

Physicist Adlène Hicheur had no idea that his life was about to be turned upside down when he joined a video conference from his home in Rio de Janeiro, Brazil, this summer to discuss his paper 'Studies of Bc + Meson decays to three-body final states at LHCb' with collaborators at CERN and elsewhere.

Police waiting downstairs whisked him to the airport, where he was summarily deported the same day. Since then, Hicheur has since found himself in a disturbing situation, detailed in a News story on page 287.

Brazilian authorities sent him to France, where Hicheur has a 2012 conviction for terrorism-related offences (and served a short prison

sentence). The French authorities placed him under house arrest, operating under sweeping detention powers given to them as part of the state of emergency declared after terrorist attacks in the country.

Leaving aside the fact that Hicheur's conviction has been vigorously contested by many scientific colleagues, a fundamental legal principle in a democracy is 'double jeopardy', which says that someone cannot be tried twice for the same offence. Yet this is effectively happening to Hicheur, both in Brazil and France. Likewise, another principle is that those who have served their sentence should have the right to pursue a new life unhindered — yet Hicheur, who by all accounts was successfully making a fresh start after moving to Brazil in 2013, and contributing productively to the country's science, has been denied this chance.

Whether one agrees or disagrees with Hicheur's house arrest — and many of his colleagues have denounced it as brutal, unjustified and unnecessary — at least it has a semblance of legal logic under the exceptional temporary situation in France.

That cannot be said of Hicheur's ejection from Brazil, which in the absence so far of a valid explanation seems to smack of arbitrariness linked to pre-Olympics tension and recent widespread coverage by Brazilian media of his past conviction. Moreover, the haste and circumstances of the action seem to violate Brazilian law, human rights and international treaties to which Brazil is a signatory.

The incident is all the more perplexing because Brazil's justice minister acknowledges that Hicheur was a law-abiding citizen during his time in the country, and France has not raised any new allegations against him. It is also difficult to reconcile the physicist described by his colleagues with the account of Hicheur in the French interior ministry's house-arrest order, which says there are "serious reasons" to think that he constitutes a security threat.

The reaction of Ignacio Bediaga, head of the group at the Brazilian Center for Physics Research in Rio de Janeiro where Hicheur first worked when he came to Brazil, echoes that of many of the deported

physicist's colleagues: "Hicheur performed an exceptional job, showed exemplary moral and ethical behaviour and a great willingness to collaborate with the group." He adds that at no time did anyone in the group perceive anything amiss with Hicheur's conduct.

Science allowed Hicheur, a Franco-Algerian citizen born in Algeria, to reach the heights of working on the Large Hadron Collider 'Beauty' experiment, better known as LHCb. After he became a *persona non grata* in European research organizations following his conviction, his international colleagues helped to find him a place to start afresh in Brazil and continue his science.

Hicheur deserves a fair and full hearing. The best route could be the Brazilian courts, and colleagues and academics there deserve support alongside Hicheur's lawyers for their efforts to pursue the case. Were Hicheur's deportation revoked, this might open the way for his return to work in Brazil, and thus make it easier for France to lift his house arrest.

In France, Hicheur is appealing his detention. But in the current climate of fear, the judicial machinery may be harder to mobilize.

French President François Hollande and his government, in their engineering of the state-of-emergency laws, have to their credit sought a difficult balance between giving police extra powers to help them fight the terrorist threat and preserving fundamental liberties and civil rights. But there is nonetheless the risk that such measures will be misused.

And if an intelligent and articulate individual such as Hicheur (a Muslim) with a bevy of support from his scientific colleagues can find himself helpless, what then of the many others with much less capacity to defend themselves? Fairness, freedom, the rule of law and human rights — including the right to a defence — are the basis for a democracy. It is not easy in these times to defend these values, much less for someone convicted in the past of terrorism-related offences, but defend them we must. ■

Bowled over

Assessing the contents of the toilet bowl in the name of crime prevention.

When they flush the toilet, most people don't think about what happens next. But for several hundred students at a private university in Washington state five years ago, what happened next was that scientists spied on some of their most intimate personal details. The researchers identified times of stress, probed the ethics of the students and calculated how many of them were bending the rules by taking drugs to help them with their degrees. The students had no knowledge of this at the time. And they probably still don't.

Likewise, the citizens of dozens of European cities have no idea that their sewage is being sifted through right now, officially to protect them; or that the police are studying the results to track crime. The toilet bowl and its contents, once extremely private, are becoming very public indeed. It's called wastewater-based epidemiology. Improved sensing techniques and analysis have made the contents of sewers and waste pipes a powerful source of data. And where there are data, there are researchers. Because although people may tell lies, the urine they send down the drain rarely does. Around for a decade or so, this analysis of waste water has mostly been used to obtain information that people would prefer others did not have — their use of illegal drugs, chiefly. Drugs broken down in the body leave telltale traces of metabolites, some of which can be found, quantified and back-calculated to work out how much of the original substance was present. Combined with a reliable estimate of the number of people who have, well, contributed a sample to the sample, the analysis can offer

guidance on average consumption and how it changes.

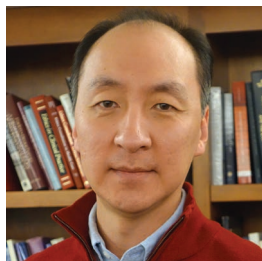
Some of the results are more worth noting than truly noteworthy. Cocaine use, unsurprisingly, peaks at the weekend. People in smaller towns and cities prefer amphetamines. And anyone watching the Netflix show *Narcos* — which chronicles the life and times of notorious drug lord Pablo Escobar — will be unsurprised to hear about the truly colossal amounts of cocaine that pass through the residents and into the waste water of the city of Medellín, Escobar's one-time heartland.

Even the study that involved the Washington students merely seemed to confirm what most people already accept: healthy university students take prescription-only medicines as 'smart drugs' to try to boost their cognitive abilities at exam time (D. A. Burgard *et al. Sci. Tot. Environ.* **450–451**, 242–249; 2013).

A paper in the journal *Forensic Science International* this month offers an intriguing new possibility. Swiss researchers describe how they hooked up with drug-enforcement investigators to use wastewater analysis to shed light on the structure of drug markets, the criminals who controlled them, and how much influence police operations had on supply (F. Been *et al. Forensic Sci. Int.* **266**, 215–221; 2016). The results are not foolproof — analysis of cannabis metabolites is chemically tricky, for example, and cannot distinguish between all sources — but the study did report some successes.

Heroin use in Lausanne was estimated by measuring morphine in the sewers and subtracting what was known to have been prescribed medically. Between October 2013 and December 2014, the scientists estimated that average daily consumption of pure heroin in the city was 13 grams. During the study, the police arrested two dealers, and analysis of phone records and interviews with users suggested that the dealers sold about 6 grams a day between them — about half the total market. This supported police intelligence that heroin, unlike other drugs such as methamphetamine, was supplied by a small number of local dealers who could be effectively targeted. You can flush, but you can't hide. ■

DANIEL MENDEZ



Illusory fears must not stifle chimaera research

Human–animal embryos have great biomedical potential — but scientists will have to quell public alarm if funding for such work is restored, says Insoo Hyun.

After more than a decade of controversy, the United States is nudging towards approving research on human–animal embryos. Last week, the National Institutes of Health (NIH) closed a month-long public consultation on ‘chimaera research’, and is widely expected to lift a moratorium that forbids federal funding for such work. Human–animal chimaeras are essentially research animals that contain transplanted human cells. Such biologically mixed animals have long been used as staple experimental systems in biomedical studies, including cancer and AIDS research. But, for some, adding human stem cells to animal embryos is a step too far — which is why the NIH imposed the moratorium, in 2015. Before then, it funded chimaeric embryo studies as long as they did not use primate blastocysts.

Chimaeric-embryo research has a vital role in basic and translational stem-cell science, so for the NIH to restore funding would be encouraging. The transfer of human stem cells into animal hosts can advance our understanding of human development and disease, and could eventually lead to the growth of transplantable human organs in livestock.

Still, the availability of federal funds does not guarantee that the research will proceed. Several states — including my own, Ohio — have raised the prospect of laws to ban such research. Institutional stem-cell review boards could still block projects, and hostile public opinion could again place future federal funds in jeopardy. Indeed, there are already signs that the NIH consultation has led to renewed protests against the research.

For these reasons, it is important for scientists to make the case for chimaera research, and to understand why opponents do not want it to proceed. Critics are especially uneasy about studies that could result in chimaeric animals with human cellular and functional modifications to the central nervous system. They argue that the transfer of human cells into animal embryos, or into the central nervous systems of animal hosts, elevates chimaeras to something approaching, or equaling, human moral status. This conflation of the biological humanization of chimaeric animals with their moral humanization is fallacious. The moral status of humans is not automatically assured by our genetic composition or the physical arrangement of our cells. Rather, it is sustained by a complex of mental traits that are fully realized only within what the Swiss philosopher Jean-Jacques Rousseau referred to as the “bosom of society”.

The moral-humanization concern distracts from what is most important in the chimaera debate. The central ethical distinction is not some ancient philosophical division between man and animal; instead, it lies in knowing the right and wrong ways to treat sentient beings according to the complexities of their attributes. The NIH has proposed that its internal steering committee could assess

chimaera-research proposals by focusing on considerations such as the characteristics of the host animal, the physical and behavioural changes likely to be caused by human-cell transfers, and incremental research monitored to determine the effects of chimaerism.

This regulatory approach is consistent with new professional guidelines for stem-cell research offered by the International Society for Stem Cell Research. Its current standards for chimaera research are based on an advisory report drafted by me and other members of its ethics committee. We urged regulators to build on animal-welfare principles in a stem-cell-specific manner, and to avoid unwarranted ‘stem-cell exceptionalism’, whereby research would be restricted by a hazy concern about the possibility of ‘morally significant’ human characteristics in chimaeric animals. The NIH and other decision-makers should heed this call.

Grounding the ethics and regulation of human–animal chimaera research in anything other than animal welfare would invite practical and philosophical difficulties. For example, one argument used against the transfer of human stem cells into early animal embryos is that this research is not overseen by animal-research committees when it is limited to experiments *in vitro*.

The challenge for these critics, then, is to explain why animal embryos containing human cells deserve serious consideration of their moral status — enough to potentially rule out their use — when standard human embryos can be used in other projects. Chimaera studies that involve sentient animals are already tightly regulated by the US Animal Welfare Act — the first federal law governing the use of animals in

research, passed 50 years ago last month — and by other national and international research policies. Under these strictures, animal-welfare principles remain the regulatory focus for all species permitted for scientific use. Because the transfer of human stem cells could have unpredicted effects on a chimaeric animal’s capacity to suffer, it is crucial that qualified veterinary staff and researchers monitor experiments for deviations from normal behaviours and species-typical functioning, and use clear criteria for humane interventional euthanasia.

The NIH’s planned approach does this, and could provide useful information on human stem cells’ possible developmental effects on animal systems, thereby aiding future oversight efforts. Such an arrangement has worked well in monitoring transgenic and knockout-animal models. It can work well for stem-cell chimaera research, too. ■

Insoo Hyun is associate professor of bioethics at Case Western Reserve University School of Medicine in Cleveland, Ohio. He is the author of *Bioethics and the Future of Stem Cell Research*. e-mail: insoo.hyun@case.edu

**HUMAN
MORAL STATUS
IS NOT ASSURED
BY OUR GENETIC
COMPOSITION
OR THE
ARRANGEMENT
OF OUR CELLS.**

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

METEOROLOGY

Air particles boost rain extremes

As the climate warms, tiny particles suspended in the atmosphere may have a greater effect than greenhouse gases on increasing the frequency of extreme rain and snowfall.

Greenhouse gases and atmospheric aerosols both drive extreme precipitation, which is expected to increase with climate change. To tease apart the climate effects, Zhili Wang of the Chinese Academy of Meteorological Sciences in Beijing and his colleagues used a global climate model to simulate scenarios with different levels of greenhouse-gas emissions.

They predict that, by the end of the century, aerosols will be two to four times more important than greenhouse gases in boosting precipitation extremes worldwide. Reducing aerosol emissions could help people to alter future climate-change impacts.

Geophys. Res. Lett. <http://doi.org/bqdf> (2016)

EVOLUTION

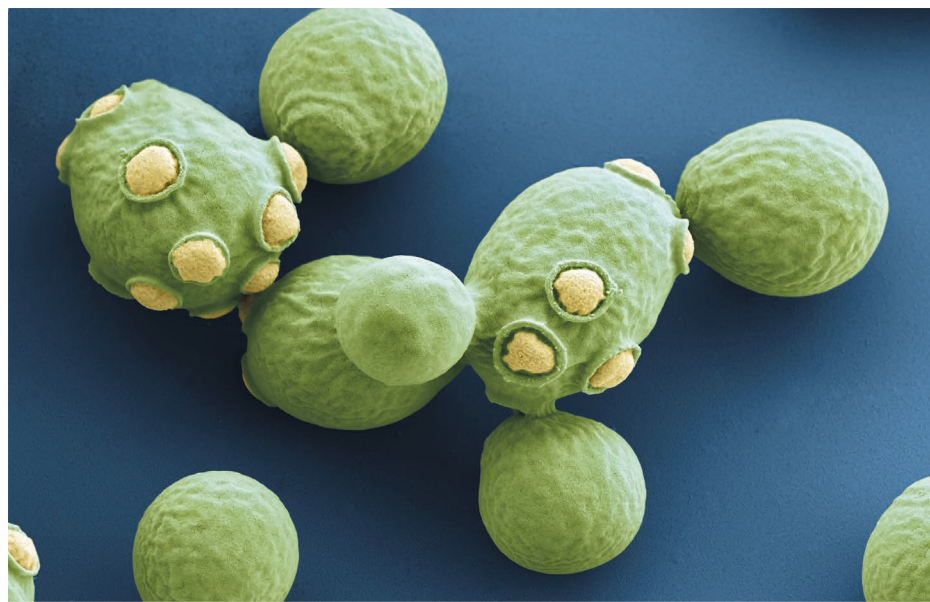
Why some groups have more species

Plants have diversified at almost twice the rate of animals, and animals and plants have accumulated new species some ten times faster than prokaryotes such as bacteria.

Across the tree of life, some groups have many more species than others. To find out why, Joshua Scholl and John Wiens at the University of Arizona in Tucson collated published data on the number of species and their phylogenetic relationships in each group of living organisms. Contrary to some hypotheses,

older groups did not have more species than young groups. Instead, the authors found that the balance of speciation and extinction over time, known as the diversification rate, determined most differences in species number between groups.

Ecological and evolutionary differences between the kingdoms of life could explain differences in diversification rates, the researchers say. *Proc. R. Soc. B* 283, 20161334 (2016)



GENOMICS

History of brewer's yeast revealed

People began to domesticate beer yeasts in the late sixteenth or early seventeenth century, when beer-making in Europe moved from homes to pubs and monasteries.

Kevin Verstrepen at the University of Leuven and Steven Maere at the University of Ghent, both in Belgium, and their colleagues sequenced the genomes of more than 150 strains of *Saccharomyces cerevisiae* (pictured) used to make bread, beer and other drinks. An evolutionary tree of the strains revealed distinct families of yeast, such as one used to make wine

and another sake, as well as two distantly related groups of ale yeast. The beer yeasts showed the strongest signatures of human influence. Beer-making strains carried variations and duplications of genes that break down maltose and maltotriose, the main sugars in beer.

The team used the genomic information to make a hybrid strain that has a high tolerance to alcohol and does not produce 4-vinyl guaiacol, which imbues unpopular clove and smoke flavours.

Cell 166, 1397–1410 (2016)

MICROBIOLOGY

Nanoparticles kill resistant bacteria

A synthetic polymer clears infections in mice caused by a multiple-drug-resistant bacterium.

Gram-negative bacteria are particularly hard to kill once they become drug resistant. To target them, Eric Reynolds, Greg Qiao and their colleagues at the University of Melbourne in Australia designed

star-shaped antimicrobial nanoparticles made of amino acids. The molecules killed several common Gram-negative pathogens in culture, and cleared infections in mice caused by *Acinetobacter baumannii*, which is resistant to several antibiotics. When cultured with sublethal concentrations of the nanoparticles for 24 days, *A. baumannii* did not grow resistant over 600 generations.

The nanoparticles hit multiple targets — disrupting

EYE OF SCIENCE/SPL

the bacterial outer membrane and the exchange of ions, and inducing pathways for cell death — and are likely to be more stable and less toxic than most antimicrobials under development, the authors say. *Nature Microbiol.* 1, 16162 (2016)

ELECTRONICS

Protection for transistors

The performance of transistors made of black phosphorus can be maintained with the addition of tellurium.

Layers of black phosphorus just a few molecules thick show great promise in advanced electronic devices. But exposure to oxygen and moisture causes damaging corrosion and bubbles to form within days. To avoid this, Zhongyuan Liu of Yanshan University in Qinhuangdao, China, and his colleagues produced samples of the material that were doped with the rare metalloid tellurium. This slowed bubble growth, and the material retained 50% of its conductivity after three weeks, whereas the undoped versions retained only 2%.

Similar approaches could allow black phosphorus to be used in high-performance batteries and computer memory, the authors say. *Adv. Mater.* <http://doi.org/f3rcsr> (2016)

ASTRONOMY

Galaxy collisions make waves fast

When galaxies with supermassive black holes at their centres collide, they could produce a burst of gravitational waves within just 10 million years.

Gravitational waves were first detected earlier this year, sparking great interest in finding more. Some scientists have predicted that wave production happens on timescales of a billion years or more, which would mean future searches would detect relatively few waves.

Fazeel Mahmood Khan at the Institute of Space Technology in Islamabad and his colleagues simulated a galaxy collision and predicted that there are many more such waves to detect.

This is a promising finding for projects that aim to look for gravitational waves, such as one proposed by the European Space Agency using the Evolved Laser Interferometer Space Antenna. *Astrophys. J.* 828, 73 (2016)

INFECTION

Feed a virus, starve a bacterium

Feeding mice helps them to fight viral infection, whereas starvation is a better strategy against bacterial infection — lending support to the proverb ‘feed a cold, starve a fever’.

Ruslan Medzhitov and his colleagues at Yale University School of Medicine in New Haven, Connecticut, studied the effects of feeding on mice that were infected with either the bacterium *Listeria monocytogenes* or an influenza virus. Bacterium-infected mice that were deprived of food stayed alive, whereas well-fed animals died. By contrast, almost all mice with flu died when they were starved, but most survived when they were fed. During bacterial inflammation, glucose from food inhibited a metabolic process that protects brain tissue from damage, whereas the sugar protected the brain during viral inflammation.

The findings suggest that different types of inflammatory response have their own metabolic programs. *Cell* 166, 1512–1525 (2016)

ENGINEERING

Fabric harvests two energy forms

A lightweight fabric can harvest both solar and mechanical energy to power electronic devices.

Zhong Lin Wang at the Georgia Institute of

Technology in Atlanta, Xing Fan at Chongqing University in China and their co-workers wove a fabric (**pictured**) using wool fibres and two types of polymer wire: a photovoltaic one and another that collects mechanical energy. The 320-micrometre-thick flexible fabric converted energy from both sunlight and movement, making enough electricity to charge a mobile phone or power a wristwatch.

Along with exploiting solar power, such a device could harvest energy from the motion of walking, the wind blowing or a moving car. *Nature Energy* <http://dx.doi.org/10.1038/nenergy.2016.138> (2016)

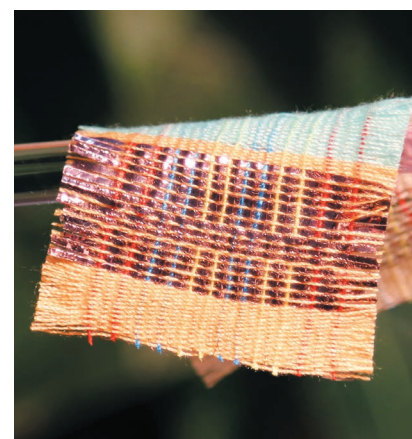
CONSERVATION

Hawaiian bird-life collapse

Populations of native birds on the Hawaiian island of Kauai have declined drastically in the face of climate change.

Eben Paxton, of the US Geological Survey's Pacific Island Ecosystems Research Center in Hawaii, and his colleagues analysed data on seven native species of forest bird on Kauai. Between 2000 and 2012, populations of six of these (including *Drepanis coccinea*; **pictured**) shrank by an average of 68% in their core range in the island's interior, and by an average of 94% in the surrounding areas. Two of these species could be detected only in the interior region in 2012 surveys.

The main driving force behind these declines



is probably increased temperatures that have allowed the spread of avian malaria, the authors say. They add that native birds are likely to go extinct in the next few decades at the current rates of decline. *Sci. Adv.* 2, e1600029 (2016)

CANCER BIOLOGY

Location matters in cancer growth

A tumour's genetic mutations often dictate which metabolic pathways it uses for rapid growth, but the tissue it develops from can also be an important factor.

Matthew Vander Heiden at the Massachusetts Institute of Technology in Cambridge and his colleagues studied tumours that bore mutations in two genes — *Kras* and *Trp53* — and that grew in either the lung or the pancreas in mice. They found that lung tumours tended to incorporate certain amino acids into proteins, and to use these amino acids as a source of nitrogen. But in the pancreas, the tumours relied less heavily on metabolizing the amino acids than the lung tumours did.

Personalized treatments for cancer should take into account both a tumour's genetics and its location, the authors say. *Science* 353, 1161–1165 (2016)

➔ **NATURE.COM**

For the latest research published by Nature visit:

www.nature.com/latestresearch



SEVEN DAYS

The news in brief

RESEARCH

DeepMind speaks

A synthesized-speech system that seems to mimic human voices more closely than ever before was revealed by Google's artificial-intelligence team DeepMind on 8 September. Rather than combining fragments of recorded speech, as current 'text-to-speech' systems do, WaveNet is a 'neural network' — a system that mimics the human brain — that is trained on raw soundwaves. It then uses statistical analysis to select which samples of audio to put together. WaveNet can also model other audio types, including music, says London-based DeepMind. Listen to WaveNet's outputs at go.nature.com/2bxil5u.

A world less wild

The world has lost 10% of its 'wilderness' areas over the past two decades, according to a report published on 8 September, with losses most acute in the Amazon and central Africa (J. E. M. Watson *et al. Curr. Biol.* <http://doi.org/bqh4>; 2016). Using satellite imagery and other data, a team led by James Watson at the University of Queensland in Australia mapped the rapid decline of wilderness areas — landscapes that are mostly biologically and ecologically intact and largely free of human disturbance. The authors argue that current conservation policies have failed to protect such areas.

Parasite honour

US President Barack Obama has another accolade to add to his Nobel Peace Prize — researchers have named a newly discovered parasite after him. *Baracktrema obamai* is a tiny flatworm that infects the blood of turtles in Malaysia



LORENZO GALASSI/AP

Surgeon guilty of scientific misconduct

Sweden's national ethics board, the CEPN, declared on 9 September that disgraced surgeon Paolo Macchiarini was guilty of scientific misconduct in a 2014 study (*Nature Commun.* 5, 3562; 2014). The paper was used to justify experimental surgeries in humans that are now deemed unethical, and describes the transplantation of a tissue-engineered oesophagus into rats at the Karolinska Institute in Stockholm. The CEPN says that the authors failed to provide them with all the raw data, and that the results in the paper were not consistent with the limited data they saw. As lead author, Macchiarini is ultimately responsible

for the content of the paper, said the CEPN's expert group on research misconduct, which investigated the study and is also looking into several other publications that he authored.

The group noted that all 23 authors shared responsibility, but expressed "some sympathy for the junior researchers", who were dependent on Macchiarini and on other team leaders. Swedish police are investigating possible charges of involuntary manslaughter and grievous bodily harm in relation to three trachea-transplant operations by Macchiarini at the Karolinska University Hospital. Two patients died and one still requires hospital care. See page 289 for more.

(J. R. Roberts *et al. J. Parasitol.* 102, 451–462; 2016). Author Thomas Pratt insists that the naming is an honour, saying of the parasite: "It's long. It's thin. And it's cool as hell."

BUSINESS

Diabetes venture

A joint venture between French drug firm Sanofi and Verily Life Sciences (formerly Google Life Sciences) of Mountain View, California, will aim to develop

better ways to treat type 2 diabetes. Onduo, to be based in Cambridge, Massachusetts, will combine Sanofi's biomedical expertise with Verily's work on devices and software to develop treatments for the disease. The firms have provided funding and capital worth about US\$500 million. The move, announced on 12 September, is another sign of life scientists' growing interest in harnessing technology firms to tackle biomedical challenges.

PEOPLE

Biology activist

Biologist and social activist Ruth Hubbard Wald — the first woman to receive tenure in biology at Harvard University in Cambridge, Massachusetts — died on 1 September, aged 92. Hubbard studied the biochemistry of vision, but became increasingly interested in politics and feminism after she got tenure in 1973. She criticized the focus

KATE WESTAWAY/GETTY on genes as a determinant of behaviour, and highlighted the dearth of research into women's health and the influence of politics on science. She co-edited an early work on gender bias in science, the 1979 book *Women Look at Biology Looking at Women*.

EVENTS

Nuclear test

North Korea said that it conducted its fifth nuclear-bomb test on 9 September, confirming seismographic detections made around the world that morning. The energy of the roughly magnitude-5 event pointed to a blast strength of at least 10 and perhaps 20 kilotonnes of TNT equivalent. This is larger than the previous tests and similar to the strength of the weapons dropped on Japan. North Korea also said that it was making progress on mounting nuclear warheads on ballistic rockets.

Happy humpbacks

Most populations of humpback whales (*Megaptera novaeangliae*; pictured) have been removed from the US endangered-species list, the US National Oceanic and Atmospheric Administration announced on 6 September. Nine of the world's 14 distinct populations are no longer in danger of extinction, thanks to global conservation efforts



over the past half-century. Commercial whaling had severely reduced the whales' numbers, and the United States listed all populations as endangered in 1970.

Renewables failure

The United Kingdom is set to miss its legally binding 2020 target to provide 15% of energy from renewable sources, according to a report from the parliamentary Energy and Climate Change Committee. A subtarget to produce 30% of electricity from renewables is likely to be met, but this will be undermined by failure to hit goals of 12% for heat and 10% for transport fuel. The report in part blames government departments that "have not cooperated effectively".

Asteroid mission

NASA's OSIRIS-REx asteroid mission launched successfully on 8 September from Cape Canaveral in Florida. The goal of the seven-year mission is to retrieve rocks and dust from

Bennu, a near-Earth asteroid, and return them to Earth. In 2020, OSIRIS-REx will descend close to Bennu's surface and deploy a 3.35-metre arm that will release a jet of nitrogen gas to loosen surface material, at least 60 grams of which should be collected. The mission, if successful, will be a proof of concept for efforts to exploit asteroids for scientific or commercial gain.

AWARDS

Biomedical prizes

Two sets of awards that are often seen as predictors for the Nobel prizes have been announced. On 13 September, the 2016 Albert Lasker Basic Medical Research Award went to William Kaelin at Harvard Medical School, Peter Ratcliffe at the University of Oxford and Gregg Semenza at Johns Hopkins University School of Medicine, for their discovery of the pathway that cells use to adapt to changes in oxygen availability. The

COMING UP

21 SEPTEMBER

The United Nations holds a meeting on antimicrobial resistance in New York City.
go.nature.com/2c9xqet

25-27 SEPTEMBER

A symposium celebrates ten years of induced pluripotent stem cells.
cell-symposia-ips.com

Lasker-DeBaakey Clinical Medical Research Award was given to Ralf Bartenschlager at Heidelberg University, Charles Rice at the Rockefeller University and Michael Sofia, chief scientific officer of Arbutus Biopharma, for their work on the hepatitis C virus. A special-achievement award went to molecular biologist Bruce Alberts, former president of the US National Academy of Sciences and former editor-in-chief of *Science*. Each award comes with US\$250,000. On 12 September, neuroscientist Reinhard Jahn of the Max Planck Institute for Biophysical Chemistry won the \$770,000 Balzan Prize for molecular and cellular neuroscience.

FACILITIES

NASA and eLISA

The director of NASA's astrophysics division said on 9 September that the agency is considering boosting its support for eLISA, a proposed space-based gravitational-wave observatory. NASA was to be an equal partner in a US\$2-billion collaboration with the European Space Agency (ESA), but in 2011, budget troubles led it to scale back its contribution. ESA officials reportedly said that an infusion of US cash and expertise could bring eLISA's target launch date forward from 2034 by several years.

NATURE.COM

For daily news updates see:
www.nature.com/news

TREND WATCH

Efforts to bring the Paris climate deal into force have been bolstered by the United States, China and Brazil formally joining this month. For the pact to take effect, 55 countries covering 55% of global carbon emissions must join; 28 nations representing 41.5% of emissions have now done so. But it's unclear what combination of countries might push the deal into effect. At least 58 nations are expected to join by the end of this year, but many are small emitters. The deal is likely to hinge on how quickly the major emitters come through.

HOW MIGHT THE PARIS DEAL COME INTO FORCE?

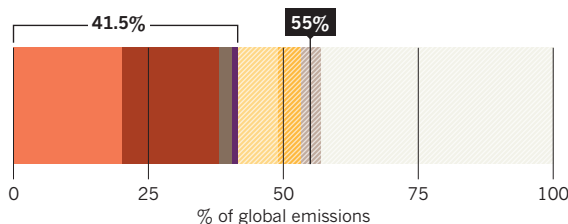
The Paris agreement needs 55 countries representing at least 55% of global emissions to join up to become legally binding.

Formally joined

China: 20% United States: 18% Brazil: 2.5% Others (25 nations): 1%

Big emitters yet to join

Russia: 7.5% Japan: 3.8% India: 4.1%



The European Union, whose 28 member states account for 12% of global emissions, is likely to join the agreement as one bloc.

NEWS IN FOCUS

MEDICINE Cancer ‘moonshot’ gets research targets amid funding uncertainty **p.288**

ETHICS How Nobel prize committee is dealing with a surgical scandal **p.289**

WILDLIFE Discovery of four giraffe species changes conservation strategy **p.290**

WORK ENVIRONMENT How to build an evidence-based healthy office space **p.294**



Adlène Hicheur at his trial in France in 2012. The physicist moved to Brazil a year later, but has now been deported to France, to the dismay of his colleagues.

POLITICS

Physicists protest against mystery deportation

Adlène Hicheur's ejection from Brazil to France remains unexplained.

BY DECLAN BUTLER

At around lunchtime on 15 July, police removed particle physicist Adlène Hicheur from his home in Rio de Janeiro and escorted him to the airport. That evening, they commanded him to board a flight to Paris, accompanied by three Brazilian police officers. From there, he was transported to his parents' home in the small southeastern town of Vienne and placed under house arrest. He must

report to police three times a day and cannot leave home between 8:00 p.m. and 6:00 a.m..

Two months later, the reasons for Hicheur's sudden deportation remain a mystery. In 2012, a French court convicted him of plotting with al-Qaeda's North African branch to carry out terror attacks on military and economic targets on French soil. Hicheur and his supporters, including scientific colleagues, maintain his innocence and say his trial was a miscarriage of justice. Brazilian authorities had discussed

his past with scientists before allowing the physicist to come to work in Brazil in 2013.

Once back in France, Hicheur was placed under house arrest using state-of-emergency powers introduced following a spate of terrorism attacks; officials there say that he still constitutes a security threat.

His colleagues, with the backing of several institutions, are ramping up their pleas to Brazilian authorities to explain the reasons for the deportation. They are concerned ►

► that it violated Brazilian law and breached Hicheur's human rights. Neither Hicheur nor his institution, the Federal University of Rio de Janeiro (UFRJ), has been given a justification for his deportation, UFRJ colleagues say, and Hicheur had no chance to contest its legality.

"His deportation without any explanation is something that makes me feel ashamed for my country," says Ron Shellard, director of the Brazilian Center for Physics Research (CBPF) in Rio de Janeiro. "If there is no objective reason for this extreme act, the Brazilian government should revoke the act of deportation and request the French authorities to send him back to Rio."

At the airport, Hicheur repeatedly requested that he be sent to Algeria (the nationality on his Brazilian work visa) or anywhere other than France, fearing that he would be confined under the state-of-emergency laws, says Ignacio Bediaga, a physicist at the CBPF. Bediaga and three UFRJ officials had rushed to the airport and remained with Hicheur until his flight took off. "In my opinion, Dr Hicheur was illegally extradited, at the request of the French government," Bediaga says.

Collaborators at CERN, Europe's particle-physics laboratory near Geneva, Switzerland, and at other European laboratories, have also expressed solidarity with Hicheur. And an international group of researchers has written to French President François Hollande, asking him to intervene to lift the physicist's house arrest — but has received no reply. Neither French nor

Brazilian authorities had responded to *Nature's* requests for comment by the time this article went to press.

Hicheur says his latest problems began in January, when the Brazilian magazine *Epoca* splashed his French conviction on its front page under the headline "A terrorist in Brazil".

"His deportation without any explanation is something that makes me feel ashamed for my country."

A deluge of media coverage followed. "I was an invited professor at the UFRJ with a smooth, peaceful life, until the craziness reached me again," Hicheur says. After Hicheur's deportation, the justice ministry issued a brief statement saying little more than that the decision was based on a recommendation by the federal police, and that Hicheur's presence was an "inconvenience to the national interest". In an interview with the newspaper *Folha de S.Paulo*, justice minister Alexandre de Moraes said Hicheur had not communicated with terrorist groups, or committed any crime while in Brazil. But he said he felt it was "absurd" to allow someone who had been convicted of terrorism-related offences to live and work in the country. "Furthermore, he is a nuclear physicist, who, in a laboratory, has all the material at hand," he added — apparently unaware that Hicheur studies the physics of fundamental particles.

But Shellard says that he discussed Hicheur's

past with Brazil's foreign office when he and others invited the physicist to Rio in 2013. Because Hicheur had served his prison term in France, and had recommendations from leading scientists, officials had no problem with his coming to Brazil.

Concern over the case is growing. On 1 September, researchers at the UFRJ's Laboratory of Elementary Particles petitioned Brazil's justice, science and education ministries to release Hicheur. The petition has now been signed by more than 300 people: mostly Brazilian physicists, but also a large contingent of researchers from European institutes. And on 5 September, a general assembly of the particle-physics section of the Brazilian Society of Physics — held at the society's annual meeting in Natal — agreed unanimously to send a letter to de Moraes, expressing concern that the society's board still hasn't received an explanation for the deportation, two months after it was first requested.

Bediaga and other researchers are convinced that repressive measures in the run-up to Rio's Olympic games, combined with media coverage of Hicheur's earlier conviction, were linked to the decision to deport the physicist.

Nadine Borges, a lawyer and human-rights expert at the UFRJ, says that she is taking up Hicheur's case in a personal capacity. In France, Hicheur's lawyers filed in July to have his house arrest lifted, but the request was quickly rejected by a Grenoble tribunal. Hicheur says he now will appeal to a higher court. ■

MEDICAL RESEARCH

Wishlist set for cancer 'moonshot'

From immunotherapies to diagnostics, experts outline research goals for US initiative.

BY HEIDI LEDFORD

Advisers to the US government's Cancer Moonshot Initiative have produced a wide-ranging laundry list of research targets — even as the project's funding remains uncertain.

The ten recommendations released on 7 September include the launch of a national clinical-trial network specifically targeted at therapies that harness the immune system, and the creation of a 3D cancer atlas to catalogue a tumour's mutations and its interactions with neighbouring normal cells.

The advisory panel — whose members include leading cancer researchers, physicians and patient advocates — also called for new technologies, including advanced imaging techniques and drug-delivery devices; a focus on proteins that drive many paediatric cancers; and studies of how tumours



Childhood cancer is a high priority for experts.

become resistant to cancer treatments.

The National Cancer Institute (NCI) has not yet determined how much funding each initiative will receive, or how the projects will be structured.

The White House launched the moonshot in January to double the pace of cancer research over the next five years. But the programme is stuck in funding limbo as Congress hashes out next year's budget. The US National Institutes of Health requested US\$680 million for the project for the 2017 fiscal year, which starts on 1 October. Despite vocal support from members of both political parties, lawmakers have said that they need more detail on the programme before they can fully fund it.

If that does not happen before Congress sets the government's 2017 budget, full funding might have to wait until fiscal year 2018, says Matt Hourihan, director of the research and development budget and policy programme at

SANDY HUFFAKER/REDUX/REX/REX

the American Association for the Advancement of Science in Washington DC.

The advisory panel's recommendations should provide the information that lawmakers want, says Jon Retzlaff, managing director of science policy and government affairs for the American Association for Cancer Research in Washington DC. Retzlaff plans to start lobbying Congress with the recommendations in hand. "The concepts and the grant proposals that will be generated because of these proposals, I think, will inspire Congress to say, 'Yes, this is a worthy project,'" he says.

For now, uncertainty hangs heavy over moonshot-related discussions. At a meeting on 7 September, NCI deputy director Dinah Singer said that the agency aims to launch some moonshot programmes in fiscal year 2017 and might seek extra funding from the private sector. But some NCI advisers are concerned that without substantial new government cash, implementing the advisory-panel recommendations could hamper the NCI's current projects.

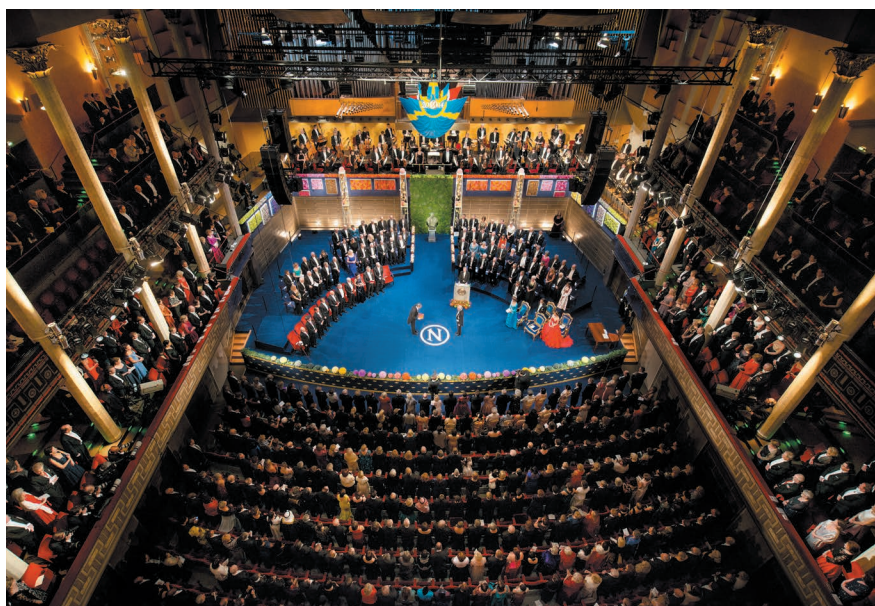
Agency director Douglas Lowy is hoping for a big budget boost from the government. "If we didn't get one, it's not that we wouldn't be able to start anything," he said. "It's just that the size, scope and speed would be dramatically different."

Despite the uncertainty, the report generated excitement among some cancer researchers. A call to expand the use of proven cancer-prevention and early-detection strategies was a pleasant surprise, says cancer geneticist Bert Vogelstein of Johns Hopkins University in Baltimore, Maryland. Although many specialists think that the approach could slash cancer deaths, it has not typically been high on the funding list, he says. "I was very impressed. They picked out some under-explored opportunities."

But at the 7 September meeting, several attendees argued that the report should have emphasized the need for research on disparities in cancer deaths that have been linked to race and economic status. "People are dying who shouldn't be dying," said Mack Roach, a radiation oncologist at the University of California, San Francisco.

That issue was largely left to the Moonshot Task Force, a separate advisory panel that is focused on improving access to cancer care and removing barriers to cancer research, said its leader, Greg Simon, chief executive of Poliwogg, a health-care investment company in New York City. The task force plans to release its report later this year.

The advisory panel's recommendations could not cover the gamut of cancer research, but the breadth of its recommendations was still impressive, says Stephen Elledge, a geneticist at Harvard Medical School in Boston, Massachusetts. "They did a pretty good job," he says. "I was glad they didn't just say, 'Oh we just need to sequence more tumours.'" ■



The medicine prize is awarded at a prestigious ceremony in Stockholm.

ETHICS

Nobel Assembly deals with scandal

Prize-selection panel rocked by investigations into surgeon — but its credibility stays intact.

BY ALISON ABBOTT

In an unprecedented move, the group that selects the winners of the Nobel Prize in Physiology or Medicine — the Nobel Assembly — has asked two of its members to resign following a scandal at the institute that supplies the assembly's members.

But scientists around the world don't see the events at the Karolinska Institute (KI) in Stockholm as a threat to the reputation of the medical prize. They say that the assembly is sufficiently separate to the KI and has handled the affair well so far.

"Everything is exploding now, but the long-term credibility won't be affected," says cancer researcher Julio Celis, associate scientific director of the Danish Cancer Society Research Center in Copenhagen.

The scandal involves the surgeon Paolo Macchiarini. Multiple inquiries have alleged that he committed scientific misconduct and subjected patients to unethical, experimental tracheal transplant operations, three of which occurred at the affiliated Karolinska University Hospital. Two of the patients have since died, and the third has required continuous hospital care since the transplant. In June,

Swedish public prosecutors opened investigations following preliminary charges against Macchiarini of involuntary manslaughter and causing grievous bodily harm. Macchiarini has denied the allegations.

On 5 September, an independent report that revealed institutional problems at the KI mentioned Nobel Assembly members Harriet Wallberg-Henriksson and Anders Hamsten — both former KI vice-chancellors — for their roles in hiring Macchiarini in 2010 and subsequently extending his contracts. (Hamsten resigned as vice-chancellor in February after acknowledging that he had misjudged Macchiarini; the KI dismissed Macchiarini in March.)

The call for Wallberg-Henriksson and Hamsten to resign came a day after the report and is a first for the 115-year-old panel, says neuroscientist Thomas Perlmann, secretary of the Nobel Committee, whose fixed-term members are elected from the more permanent assembly.

"The professionalism of some of the faculty at the Karolinska Institute has been called into question, and this won't go away," says Erwin Neher of the Max Planck Institute for Biophysical Chemistry in Göttingen, ▶

► Germany, who won the medicine prize in 1991. “But I don’t think this discredits the Nobel prize — they are two different things.”

When Alfred Nobel died in 1896, he left the bulk of his fortune — amassed from his explosives businesses — to the Nobel prizes. His will specified which institutions would select each prize, and declared the KI in charge of medicine. The first prizes were awarded in 1901.

At first, the entire KI faculty selected the medicine winners, but by the 1970s it had grown too large for this to be practical — and a new law made all documents at state institutions accessible to the public, ruling out secret deliberations. So in 1977, the Nobel Assembly was created, comprising 50 KI professors; the Nobel Foundation pays for its operations.

The Nobel Committee has also done a good job of separating itself from the Macchiarini affair since it began, says neuroscientist Eero Castrén at the University of Helsinki. KI geneticist Urban Lendahl, who participated in the decision to hire Macchiarini, resigned his position as secretary-general of the Nobel Committee in February, notes Castrén. (Lendahl

stepped down because he anticipated that he would be involved in the investigation.)

Two other assembly members — clinical immunologist Katarina Le Blanc, who co-authored a paper with Macchiarini that is under investigation by the Central Ethical Review Board, and Hans-Gustaf Ljunggren, who was dean of research at the KI from 2013 until February — have not been asked to resign because there is still “uncertainty over their roles” in the Macchiarini affair, says Perlmann.

“To protect the brand”, he adds, none of the three, nor Wallberg-Henriksson, nor Hamsten, has participated in assembly activities since February. Perlmann says that the Nobel Committee is not taking further action, but will monitor perceptions of the prize to see whether it needs to do more.

“It is important that institutions deal in a fair way with those whose judgement or moral probity has been called into question,” says Steven Hyman, director of the Stanley Center for Psychiatric Research at the Broad Institute in Boston, Massachusetts, who has nominated prize candidates to the Nobel Committee. “The

Nobel Assembly seems to be doing this.”

He adds: “There is no benefit to the world, or to patients who have been harmed, by using a very serious incident to undercut a globally important institute.”

The assembly has survived other challenges, usually relating to complaints about its choices. In 1994, it encountered accusations — quickly discredited — that it had allowed a drug company to buy the 1986 medicine prize for Italian neuroscientist Rita Levi-Montalcini.

Just as the Swedish king never comments on politics, the Nobel Assembly never comments on such complaints. But during its 100th anniversary celebrations, it acknowledged some regrets — such as awarding a share of the 1923 prize for the discovery of insulin to John Macleod, whose role is now questioned, and the failure to recognize Oswald Avery, who identified DNA as the genetic material in the 1940s.

“The prize has survived many things,” says cell biologist Måns Ehrenberg of Uppsala University, who has served on the committee that selects the Nobel Prize in Chemistry. “The standard of evaluation no one can criticize.” ■

BIOLOGY

DNA reveals four giraffe species

Finding could guide efforts to conserve the iconic animals.

BY CHRIS WOOLSTON

One of the most iconic African animals has a secret. A genetic analysis suggests that the giraffe is not one species, but four — a finding that could alter how conservationists protect the animals.

Researchers previously split giraffes into several subspecies on the basis of their coat patterns and where they lived. Closer inspection of their genes, however, reveals that giraffes should actually be divided into four distinct lineages that don’t interbreed in the wild, scientists reported on 8 September in *Current Biology*¹. Previous genetic studies² have found discrete giraffe populations that rarely intermingled, but this is the first to detect species-level differences, says lead author Axel Janke, a geneticist at Goethe University in Frankfurt, Germany.

“It was an amazing finding,” he says. He notes that giraffes are highly mobile, wide-ranging animals that would have many

chances to interbreed in the wild, if they were so inclined. “The million-dollar question is what kept them apart in the past,” Janke speculates that rivers or other physical barriers kept populations separate long enough for new species to arise.

RUMINATING ON RUMINANTS

The study tracked the distribution of 7 specific gene sequences — chosen to measure genetic diversity — in nuclear DNA from skin biopsies of 190 giraffes. It also analysed the animals’ mitochondrial DNA. The sequences fell into four distinct patterns that strongly suggested separate species. Janke says that the four species are about as different from each other as the brown bear

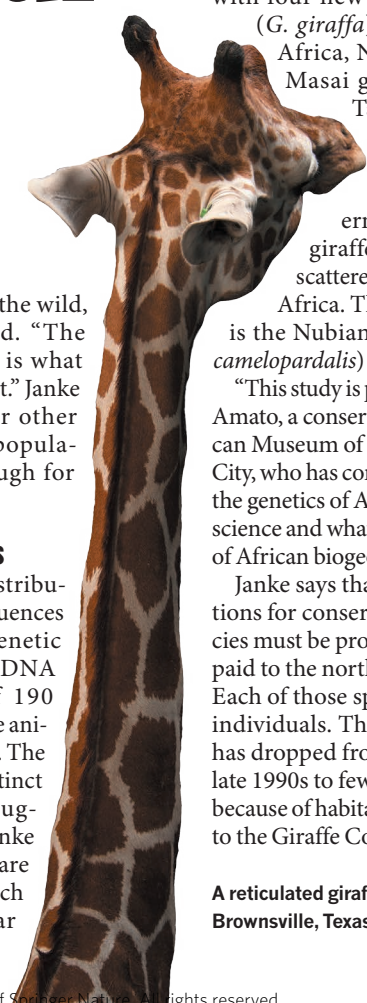
(*Ursus arctos*) is from the polar bear (*Ursus maritimus*).

The researchers suggest replacing the current species name, *Giraffa camelopardalis*, with four new ones: the southern giraffe (*G. giraffa*), found throughout South Africa, Namibia and Botswana; the Masai giraffe (*G. tippelskirchi*) of Tanzania, Kenya and Zambia; the reticulated giraffe (*G. reticulata*) found in Kenya, Somalia and southern Ethiopia; and the northern giraffe (*G. camelopardalis*), found scattered through central and eastern Africa. The one remaining subspecies is the Nubian giraffe (*G. camelopardalis camelopardalis*) of Ethiopia and South Sudan.

“This study is pretty persuasive,” says George Amato, a conservation biologist at the American Museum of Natural History in New York City, who has conducted extensive research on the genetics of African wildlife. “I applaud the science and what it adds to our understanding of African biogeography.”

Janke says that the findings have implications for conservation: all of the giraffe species must be protected, with special attention paid to the northern and reticulated giraffes. Each of those species has fewer than 10,000 individuals. The overall number of giraffes has dropped from more than 140,000 in the late 1990s to fewer than 80,000 today, largely because of habitat loss and hunting, according to the Giraffe Conservation Foundation.

A reticulated giraffe at the Gladys Porter Zoo in Brownsville, Texas.



JOEL SARTORE/NGC/GETTY

But applying the findings to conservation efforts may be difficult, because it's not always obvious how that knowledge should guide decisions about animal protection. "So far, we haven't really been able to fully appreciate the power of genomics in conservation," says Aaron Shafer, a geneticist at Trent University in Peterborough, Canada.

FINDING CLARITY

Amato notes strong parallels between giraffes and African elephants, which were classified as a single species until a 2010 study³ provided genetic evidence that there were actually two: forest elephants (*Loxodonta cyclotis*) and savannah elephants (*Loxodonta africana*). That finding increased calls for extra protection of the forest elephant, the rarer of the two.

However, assessments by the International Union for Conservation of Nature still treat the animals as one species, owing to concerns that splitting them into two would place elephant hybrids into a kind of conservation limbo.

Evidence showing that many populations of American bison (*Bison bison*) carry a little domestic-cattle DNA⁴ prompted concerns over whether it was worth saving the contaminated herds, because they weren't completely wild. Amato and other biologists have argued that the animals still deserve protection. "They are ecologically functional bison," he says.

It is unclear whether this study will have any impact on giraffe conservation, says Amato. The most immediate effects may be felt in zoos that trade the mammals for breeding purposes: now that researchers have identified several species, it should be easier for zookeepers to make appropriate matches.

The discovery of these giraffe species could have come sooner, but science has neglected the animals. "Giraffes were fairly ubiquitous in their habitat, and they weren't much of a target for poachers," Amato says. "They are an iconic animal, but they were taken for granted." ■

1. Fennessy, J. *et al. Curr. Biol.* <http://dx.doi.org/10.1016/j.cub.2016.07.036> (2016).
2. Brown, D. M. *et al. BMC Biol.* **5**, 57 (2007).
3. Rohland, N. *et al. PLoS Biol.* **8**, e1000564 (2010).
4. Hedrick, P. W. *J. Hered.* **100**, 411–420 (2009).



Studies of people who survived Ebola are altering scientists' understanding of the virus.

PUBLIC HEALTH

Ebola virus and its legacy linger on

Long-term tracking of people who beat the virus reveals its remarkable longevity in the human body.

BY ERIKA CHECK HAYDEN

Ebola survivors are teaching scientists some surprising lessons. Long-term studies have revealed that the virus lasts longer in survivors' bodies than previously suspected.

The findings, presented on 12 September at an Ebola-virus conference in Antwerp, Belgium, underscore the need for extended tracking of people who have beaten Ebola and other rare infections. Researchers have long known that the virus can persist in people who have recovered from the infection. But the size of the West African outbreak, coupled with improved monitoring technologies, is changing how scientists view life after Ebola — and how to prevent future outbreaks.

"Now that you have tens of thousands of survivors and systemic approaches to follow them, you can detect things that happen more rarely and attribute them to Ebola," says physician and epidemiologist Daniel Bausch of the World Health Organization in Geneva, Switzerland.

Researchers will soon publish the first confirmed report of a person without obvious Ebola symptoms infecting another person. A seemingly healthy mother in Guinea passed the virus to her nine-month-old daughter in breast milk, and the child died from Ebola-virus infection in August 2015, according to a European Union-funded team led by Sophie Duraffour from the Bernhard Nocht Institute for Tropical Medicine in Hamburg, Germany.

A study due to be presented at the Antwerp meeting also suggests that some people ►



TOP NEWS



Ale genomics: how humans tamed beer yeast
go.nature.com/2csnuzd

MORE NEWS

- Moon's pull can trigger big earthquakes go.nature.com/2cjsv0k
- Reactor shutdown threatens world's medical-isotope supply go.nature.com/2cbfm8v
- Virus pumps up male muscles — in mice go.nature.com/2ckullw

NATURE PODCAST



The ideal office environment; synthesizing speech; and embryo epigenetics nature.com/nature/podcast

► who became infected during the recent outbreak escaped detection. Miles Carroll, an epidemiologist at Public Health England in Porton Down, and his colleagues tracked 80 people who had contact with Ebola patients in Guinea but did not themselves become noticeably ill. Yet 15–20% of these contacts developed immune responses capable of neutralizing Ebola viruses, suggesting that they had contracted mild infections that went undetected.

This 'sub-symptomatic' or 'asymptomatic' Ebola was known to exist, but the latest studies involve more people who have been studied more intensively than in the past. Researchers caution, however, that it is still rare for Ebola lingering in a person's body to spark new outbreaks. The phenomenon would probably have escaped notice if the recent epidemic had been smaller.

Thousands of men who are infected have survived, but until recently scientists did not know that the Ebola virus could be transmitted in semen beyond three months, says Mary Choi, an epidemiologist at the US Centers for Disease Control and Prevention. The agency and the Liberian government are running the largest-ever investigation of Ebola viruses in the semen of survivors. So far, the team's study of 466 men has detected virus fragments in semen up to 18 months after a man has recovered from his infection¹.

In February, two months after the outbreak was declared over in Guinea, Duraffour and her colleagues traced a cluster of new Ebola cases to a man who transmitted the virus to a sexual partner 17 months after recovering from his infection². Yet another study, which examined 26 male Ebola survivors, found that the vast majority eliminated the virus from their semen within 4 months of recovery³. The precise timing varied widely from person to person, however.

Choi says that the virus probably lasts for longer than 18 months in semen. Her team will continue to monitor the virus's persistence, while counselling survivors to use condoms or abstain from sex until their semen tests negative twice. "The primary takeaway is that semen testing should be incorporated earlier on as part of services that survivors receive," Choi says.

Researchers must show sensitivity in communicating such findings, says virologist Stephan Günther of the Bernhard Nocht Institute, and take care not to make life more difficult than it already is for Ebola survivors, who face discrimination and lingering health problems. "We have to be careful to stress that these are very, very rare events." ■

1. Soka, M. J. *et al. Lancet Glob. Health* [http://dx.doi.org/10.1016/S2214-109X\(16\)30175-9](http://dx.doi.org/10.1016/S2214-109X(16)30175-9) (2016).

2. Diallo, B. *et al. Clin. Infect. Dis.* <http://dx.doi.org/10.1093/cid/ciw601> (2016).

3. Sissoko, D. *et al. Lancet Glob. Health* (in the press).



The Gaia telescope's data will help to measure the distances to 'standard candles' such as RS Pup (centre).

GAIA SPACECRAFT

Galaxy map will change astronomy

Data will shed light on exoplanets, cosmology and more.

BY DAVIDE CASTELVECCHI

Astronomers the world over are about to get their first taste of a transformative tool. As *Nature* went to press, Gaia, a space telescope launched by the European Space Agency (ESA), was due to release its first map of the Milky Way on 14 September. Initially, the catalogue will show the 3D positions of 2,057,050 stars and other objects, and how they have changed over two decades. Eventually, it will contain one billion objects or more.

The release is expected to include 19 papers by the Gaia astronomers who have seen the data. Independent teams could produce 100 or so papers just in the weeks following the release of the draft catalogue, says Lennart Lindegren, an astronomer at the Lund Observatory in Sweden and a driving force for Gaia.

"Gaia is going to revolutionize what we know about stars and the Galaxy," says David Hogg, an astronomer at New York University. He and others are leading 'Gaia hacking' events that will attempt to exploit the burst of data. So, what might some of the discoveries be?

MILKY WAY ARCHAEOLOGY

Gaia's 3D view will reveal how stars move under the Milky Way's combined gravitational pull. This will add to knowledge of the Galaxy's structure, including that of parts not directly visible from Earth, such as the 'bars' that join the Galactic Centre to the Milky Way's spiral arms.

Researchers will also be able to identify outlying stars that stream together at high speeds and are thought to be remnants of mergers with smaller galaxies, says Michael Perryman, a former senior scientist for Gaia at ESA. Combined with data about stars' colour, temperature and chemical composition, this will enable researchers to reconstruct the Galaxy's 'archaeology' from the past 13 billion years.

WHERE IS THE GALAXY'S DARK MATTER?

The details of star trajectories will uncover the Milky Way's distribution of dark matter, which constitutes the bulk of matter in the Universe. That could help to reveal what dark matter is.

Gaia might also put some exotic theories to the test. MOND (modified Newtonian dynamics) predicts a different Galactic gravitational field from standard dark-matter theory; star

velocities measured by Gaia will be able to investigate which is right. The probe might even help to reveal whether dark matter killed the dinosaurs, as suggested by a theory from 2013.

DISPUTED STELLAR DISTANCES

Gaia will provide precise measurements of how far individual stars lie from the Sun.

One of the first groups of stars that researchers want to check is the Pleiades, a cluster in the constellation Taurus. Most observations, including one made with the Hubble Space Telescope, put the cluster about 135 parsecs (440 light years) away (D. R. Soderblom *et al. Astron. J.* **129**, 1616–1624; 2005). But results based on data from Hipparcos, an ESA space mission that preceded Gaia, suggest that it is only 120 parsecs away (F. Van Leeuwen *Astron. Astrophys.* **497**, 209–242; 2009).

The discrepancy cast some doubt on the Hipparcos result. Gaia uses a method that is similar to, but much more evolved than, that of the earlier mission, so astronomers will be watching it closely.

NEW WORLDS

Astronomers have discovered thousands of planets orbiting other stars, mostly by detecting tiny dips in a star's brightness when an orbiting planet passes in front of it. Gaia will instead seek planets by looking for slight wobbles in the star's

position caused by a planet's gravitational pull.

Gaia's technique is best suited to detecting large planets in relatively wide orbits, says Alessandro Sozzetti, a Gaia researcher at the Astrophysical Observatory of Turin in Italy. And unlike the more common transit method, it directly measures a planet's mass. If it works, it will be a striking comeback for a technique that has seen many false starts. But it will require several years of observation, with a sneak preview expected by 2018, Sozzetti says.

HOW FAST IS THE UNIVERSE EXPANDING?

Gaia explores the Milky Way, but its influence extends to the wider observable Universe.

To estimate the distances to faraway galaxies, astronomers typically use stellar explosions called Type Ia supernovae. The explosions' apparent brightnesses reveal how far away they and their galaxies are. Such 'standard candles' have been the main tool for estimating the rate of expansion of the Universe, and have led astronomers to propose that a mysterious 'dark energy' is accelerating the expansion.

The method depends on a comparison with other types of standard candle in the Milky Way. In its first release, Gaia will measure the distances to thousands of such stars. Such measurements may eventually resolve conflicting estimates of the rate of cosmic expansion.

INVISIBLE ASTEROID THREATS

As it scans the sky, Gaia is expected to discover hundreds of asteroids inside the Solar System, says Gaia astronomer Paolo Tanga of the Côte d'Azur Observatory in Nice, France.

When it spots a near-Earth object, an asteroid whose orbit brings it within about 200 million kilometres of Earth, Gaia can alert observatories to use ground-based telescopes to establish whether the object is a threat.

It will scan nearly the entire sky and might reveal objects that, during certain times, are too close to the Sun to observe from Earth, says Anthony Brown, an astronomer at the Leiden Observatory in the Netherlands and chair of Gaia's data-processing collaboration. Asteroid paths will also enable Gaia to perform sensitive tests of the general theory of relativity. ■

Read a longer version at go.nature.com/2cy81uy

CORRECTION

The News story 'Mars contamination fear could divert Curiosity rover' (*Nature* **537**, 145–146; 2016) should have made it clear that the dark streaks near Curiosity are only 'potential' recurring slope lineae. And it should have said that the Murray formation — not the Murray Buttes — was formed from ancient lake sediments.



The Well Living Lab tracks workers' responses to the indoor environment.

THE OFFICE EXPERIMENT

CAN SCIENCE BUILD THE PERFECT WORKSPACE?

BY EMILY ANTHES

In late May, eight employees of Mayo Clinic's medical-records department packed up their belongings, powered down their computers and moved into a brand new office space in the heart of Rochester, Minnesota. There, they made themselves at home — hanging up Walt Disney World calendars, arranging their framed dog photos and settling back into the daily rhythms of office life.

Then, researchers started messing with them. They cranked the thermostat up — and then down. They changed the colour temperature of the overhead lights and the tint of the large, glass windows. They played irritating office sounds through speakers embedded in the ceilings: a ringing phone, the clack of computer keys, a male voice saying, “medical records”, as if answering the phone.

On a warm morning in June, the recording is playing on a loop. “I’ve timed it,” says Randy

Mouchka, one of the relocated office workers, with exasperation. “It’s 55 seconds.” Today, the air feels stale and stuffy, but the sun is streaming in — an improvement over last week, Mouchka says, when the researchers kept the window shades pulled all the way down.

These people are the first guinea pigs in the Well Living Lab, an immersive, high-tech facility where Big Brother meets big data. The lab — a collaboration between Mayo Clinic in Rochester and Delos, a design and technology firm based in New York City — was built to host studies on how the indoor environment influences health, well-being and performance, from stress to sleep quality, physical fitness to productivity.

Down the hall, in a glass-walled control centre crammed with computers, scientists are keeping a close eye on Mouchka and his colleagues. “We have a panoramic view of

everything that’s happening,” says Alfred Anderson, the lab’s director of technology. One monitor features a live video feed; others display light levels, air temperature, humidity and atmospheric pressure from the 100 or so sensors scattered around the office. The workers are wired up, too: a large monitor reveals the readouts from biometric wristbands that measure their heart-rate variability and the electrical conductance of their skin, both crude measures of stress. Researchers will monitor all of this as they subject the employees to nine different types of office environment. “We’re in ‘Bad Office 2’ today,” Anderson says.

Experts know that indoor spaces can pose health risks. Excessive noise is thought to contribute to high blood pressure and heart disease. Artificial light can disrupt circadian rhythms and may increase the risk of certain cancers. There is growing evidence that a sedentary lifestyle could damage health, leading to type 2 diabetes, cardiovascular disease, cancer or early death — a major concern when so many modern jobs demand sitting at a desk all day. And workplace stress is thought to cost hundreds of billions of dollars worldwide each year in sick days, health-care costs and lost productivity. “We spend 90% of our time indoors,” says Brent Bauer, the Well Living Lab’s medical director. “If we don’t optimize that, we’re going to have a hard time optimizing wellness as a whole.”

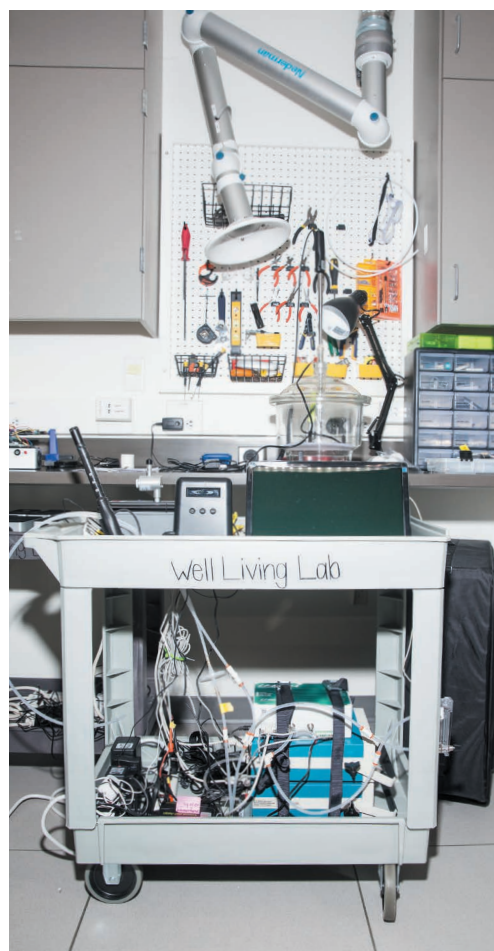
ACKERMAN + GRUBER FOR NATURE

CORRECTION

The News story 'Nobel Assembly deals with scandal' (*Nature* **537**, 289–290; 2016) erroneously gave Stockholm as the location for all of the Nobel prize ceremonies.



In the Well Living Lab's control room (above), researchers track dozens of variables — such as lighting, temperature, humidity and noise levels — using dozens of environmental sensors (left) placed throughout the office. With tools in the Hardware Development Lab (right), it can reconfigure the space into apartments, hotel rooms and more.



Scientists hope that the lab will allow them to add to the growing literature on the impact of the built environment, and to produce practical, evidence-based recommendations for creating healthier indoor spaces ranging from offices to homes. It's an ambitious mission that will involve integrating and interpreting vast quantities of data. But scientists, companies and organizations — impressed by the lab's size, scope and approach — are eager to see what it finds. "Everybody I've talked to who has heard about it is very excited because it is truly unique," says Gail Brager, associate director of the Center for the Built Environment at the University of California, Berkeley.

LIVING IN THE LAB

Decades of research have revealed that indoor spaces can affect how people think, feel and behave. In a landmark 1984 study¹, Roger Ulrich, a pioneer in health-care design research now at Chalmers University of Technology in Gothenburg, Sweden, found that people recovering from surgery in hospital rooms with views of nature needed shorter stays and fewer doses of strong pain medication than did those in rooms looking onto a brick wall. Others have reported that certain kinds of artificial light can improve sleep and reduce depression and agitation in people with Alzheimer's disease²; that higher air temperatures seem to curb calorie

consumption³; that employees take more sick leave when they work in open-plan offices⁴; and that children in daylight-drenched classrooms progress faster in maths and reading than do those in darker ones⁵.

In 2012, the accumulating research led Delos — which aims to create spaces that boost health and wellness — to start developing evidence-based guidelines for healthier buildings. The WELL Building Standard, first released in 2014, outlines more than 100 best practices, from using paints that release minimal levels of potentially toxic compounds to organizing cafeterias so that they prominently display fruit and vegetables. Buildings that meet enough of the standards can become 'WELL Certified', in much the same way that buildings can earn sustainable, eco-friendly certification.

But in developing the standard, Delos noticed gaps in the scientific literature. There were many studies on a single aspect of the indoor environment, such as light or sound, but in the real world, these variables operate in concert. Studies have shown, for example, that as the temperature and humidity of indoor air increases, its perceived quality declines⁶. Programmes to reduce indoor air pollution could yield greater benefits if building managers pay attention to these other factors.

Other recommended practices might conflict. In June, researchers reported⁷ that office

workers scored higher on tests of cognitive function when the room was better ventilated, but many studies have found that background noise impairs cognitive performance. What if increasing air flow requires office workers to open a window onto a loud street? If one worker wants quiet, and another wants fresh air, can evidence decide who should win?

"There are some building-science labs out there who try to bring in as many components as possible, but we never thought they got to the point where they really could address all the issues that might come up in a building design standard," says Dana Pillai, president of Delos's research division and executive director of the Well Living Lab. "So we thought we'll just do it ourselves." In 2013, Delos began discussions with Mayo Clinic. Together, the organizations decided to build an adaptable, immersive lab that gave them precise control over many environmental variables and mirrored the real world as closely as possible.

They assembled an 18-person team and sketched out a 700-square-metre dream lab. The facility, which cost more than US\$5 million to build and occupies the third floor of an office building, is endlessly transmutable. The tint of the windows can be altered with a mobile app; LED lighting can be tuned to different colours and intensities, and the motorized shades can be programmed to rise and fall at specific

times of day. “We can move walls, we can move plumbing, we can move ducts,” says Bauer. Researchers can transform the lab from a large, open-plan office to a cluster of 6 apartments or 12 hotel rooms, where study participants might live for weeks or even months. “It’s imaginative,” says Alexi Marmot, an architect and researcher at University College London. “This really has potential to allow all sorts of things to be done that we have not been able to do.”

The Well Living Lab occupies a scientific sweet spot — more controlled than the real offices used for field studies and more realistic than many laboratories. “That they’re going to have people there for extended periods of time, I think is really important,” says Brager, who was not involved in the planning or design of the lab, but will serve on its scientific advisory board. “While this still isn’t quite a real building — so there’s still going to be some question about the ability to generalize to real-world conditions — it’s a lot closer than the conventional labs.”

OFFICE SPACE

The Well Living Lab’s scientists are starting small and simple, drawing on previous findings to create a variety of office environments that they hypothesize will have positive, negative or no effects on workers’ comfort and stress. They are monitoring participants’ responses to these changing conditions with daily surveys — which ask for ratings of comfort, satisfaction, productivity and stress — and the biometric wristbands. This study is a trial run, designed to validate the lab’s systems and approach, as well as the basic idea that office conditions influence employees’ well-being.

Later this year, the team will explore in more detail how light, noise and temperature affect employee performance, as measured by tests of executive function and productivity, surveys of perceived productivity and physiological measures. Crucially, the researchers will also assess how variables interact, which have the greatest impact on individual and group performance, and what the cumulative effects of changing them are. Such studies might eventually show, for example, that an office with plenty of natural light, a thermostat set to 21 °C and a modest hum of background noise produces the happiest employees, who respond to e-mails quickly or enter database information accurately.

“The world is a multicomponent place, so there’s a benefit of doing that — that’s how the real world is,” says Mariana Figueiro, who directs the Light and Health Program at Rensselaer Polytechnic Institute in Troy, New York. But there’s a danger, too, she says. “Those are probably going to be very expensive studies, and they might be very noisy” statistically, which may make the data difficult to interpret.

Even the relatively simple pilot study is already generating nearly 9 gigabytes of data per week. As the researchers enrol bigger groups and monitor more variables and outcomes, that figure could expand tenfold.

The complexity will also grow as the team begins to layer studies on top of one another. Nicholas Clements, a director at Delos Labs, is collecting samples of the office microbiome: bacteria, fungi and more that live in the office’s nooks and crannies, and on the surfaces that people touch every day. Scientists think that it may be possible to actively shape the indoor microbiome to improve human health, but research into this idea is in its infancy.

“We’d like to push that science further and hopefully we can accomplish that here,” says

“WE’RE TAKING KIND OF A KID-IN-A-CANDY-STORE APPROACH.”

Clements, who plans to test whether certain environmental interventions, such as changing flooring and surface materials or installing a ‘green wall’ of living plants, can alter the office’s microbes — or the health of its human occupants. (He will also track participants’ exposure to indoor air pollutants, such as the volatile organic compounds emitted by paint and furniture.)

Other Mayo faculty members are eager to use the facility. Early next year, ergonomist Susan Hallbeck will investigate whether standing desks improve health in workers with and without certain risk factors for disease — and, if so, what the optimal ratio and schedule of standing and sitting is. Research has shown that using a standing desk can slightly increase the number of calories burnt, but the evidence for broader health benefits is limited. “This is a dream study,” says Hallbeck.

In addition to the office space, the lab currently contains a single studio apartment, which the researchers will use to learn how to design living spaces that improve sleep quantity and quality in night-shift workers, and whether changes in these workers’ circadian cycles influence their microbiota.

And whenever the scientists get together, they start churning out new ideas and hypotheses. Perhaps they could turn the space into a classroom, study whether lighting can reduce falls among older people or probe whether certain office conditions make it easier for people with traumatic brain injuries to return to work.

“We’re taking kind of a kid-in-a-candy-store approach,” Bauer says. “We’ve got almost endless opportunities now to start answering these important questions about, ‘How do we optimize the indoor environment?’”

COMPLEX CHALLENGES

The lab’s leaders still have a long wish list of sensors and technologies that they would like to deploy, and they’re eyeing international expansion. They’re not alone. A handful

of other teams are taking an immersive, multivariable approach to studying human responses to indoor conditions, using flexible facilities — from the Total Indoor Environmental Quality Lab at Syracuse University in New York, to the SenseLab at the Delft University of Technology in the Netherlands, which should open in December.

But big ambitions can be expensive. To help cover costs, Mayo and Delos have been recruiting corporations and other organizations to the Well Living Lab Alliance. Members make contributions ranging from \$75,000 to \$300,000, and receive several benefits in return, including early access to research findings, attendance at an annual Well Living Lab summit and discounts on sponsored research. So far, nine organizations — in industries including construction, property management, health-care technology, manufacturing and computing — have signed up.

Corporate partnerships aren’t unusual in built-environment research, but scientists say that the lab will have to select its members carefully, be transparent about funding sources and work to ensure scientific independence. “In this field that’s normally been neglected, there’s now somebody who clearly has very deep pockets,” Marmot says. “I think it’s all to the good. But let’s make sure that the appropriate scientific review processes are there.”

Bauer says that all proposed studies — including those sponsored by alliance members — will need approval from the lab’s leaders, its joint steering committee and Mayo’s institutional review board. “I think we’ve been very clear with the companies that are participating that membership isn’t a *carte blanche*,” he says.

At the Well Living Lab, the workers are now feeling at home. Despite being poked, prodded and observed by the scientists behind the glass, the first test participants love their temporary office. The desks are adjustable, the chairs comfy and the windows big. Even the air, they say, seems cleaner than in their old offices, to which they will eventually return. “I don’t want to go back,” says Mouchka. “I’m hoping we’re here for a year.” ■

Emily Anthes is a science journalist based in New York City.

1. Ulrich, R. S. *Science* **224**, 420–421 (1984).
2. Figueiro, M. G. *et al. Clin. Interv. Aging* **9**, 1527–1537 (2014).
3. Bernhardt, M. C., Li, P., Allison, D. B. & Gohlke, J. M. *Front. Nutr.* **2**, 20 (2015).
4. Pejtersen, J. H., Feveile, H., Christensen, K. B. & Burr, H. *Scand. J. Work Environ. Health* **37**, 376–382 (2011).
5. Heschong Mahone Group *Daylighting in Schools: An Investigation into the Relationship Between Daylighting and Human Performance* (Pacific Gas & Electric Co., 1999); available at go.nature.com/2cdfmrq
6. Fang, L., Wyon, D. P., Clausen, G. & Fanger, P. O. *Indoor Air* **14** (Suppl. 7), 74–81 (2004).
7. Allen, J. G. *et al. Environ. Health Perspect.* **124**, 805–812 (2016).

Guardian of a hidden universe

Diana Wall has built a career on overturning assumptions about soil and underground ecosystems. Now she is seeking to protect this endangered world.

BY RACHEL CERNANSKY



Diana Wall wields a coring tool in the grasslands of northern Colorado.

Early on a cold spring morning, Diana Wall is trying out a tool normally used to make holes on golf courses — and she can't contain her excitement. Her team has always used more laborious methods to take samples of soil and its resident organisms. "Oh, that's a beautiful core," she says as one student bags a sample filled with tiny roundworms. "Hello, nematodes!"

Wall, a soil ecologist and environmental scientist at Colorado State University in Fort Collins, has come to this site about an hour east of the campus to collect data for one of her latest experiments. She and her colleagues are creating an artificial drought in a patch of grassland by covering it with temporary shelters. They expect that predatory nematodes will die or enter a type of suspended animation, leaving the parasitic nematodes that prey on plants to dominate the ecosystem. "How do plants respond below-ground to drought?" she wonders.

Wall has been asking — and answering — similar questions about soil for decades. She has become one of the most celebrated and outspoken experts on the hidden biodiversity in dirt, having studied soils and their inhabitants in nearly every corner of the world. She has a special fondness for Antarctica, which she has visited almost every year since 1989. It was there that she and a colleague made a landmark discovery, demonstrating that the soil in one of the driest spots on Earth is home to some animal life and not sterile, as many had thought.

The same drive to challenge orthodoxy also helped her to advance in a field in which women were once rare. "Many times, I felt like I was hitting the glass ceiling and got discouraged," she says, before emphasizing how things have improved. "Today, I love seeing so many women in Antarctic and other research."

Alongside her own experiments, Wall has become an ambassador for soil science and conservation — at a time when soil ecosystems are being devastated by forces such as erosion, pollution, pesticides and climate change. Soil degradation over the past two centuries or so has released billions of tonnes of stored carbon into the atmosphere, and this discharge could accelerate, speeding up climate change. Beyond that, says Wall, the threats to soil could jeopardize food production, water quality and the health of humans, plants and animals. The current path, she says, "leaves our terrestrial biodiverse world as we know it very uncertain".

The efforts of Wall and other scientists to raise the profile of soils have been making an impact. The United Nations declared 2015 the International Year of Soils, and in May, Wall travelled to Nairobi to launch the *Global Soil Biodiversity Atlas* — a compendium of information developed by a team of more than 100 scientists, which she helped to lead.

David Montgomery, a geomorphologist at

the University of Washington in Seattle, says that Wall has inspired many other researchers in their science and outreach on topics important to society. "We need more first-rate scientists willing to speak in those arenas."

WEDDED TO THE ICE

This month, Wall is busy planning for her next trip to Antarctica, which will come, as usual, just after Christmas. Her colleagues joke that those journeys keep Wall young because she often crosses the International Date Line on her birthday, essentially erasing the day from the calendar. Assuming that she passes her physical — for which she is swimming and cycling — this trip will be her 27th to Antarctica.

Wall is 72 and has seemingly boundless energy. Tall and thin, she speaks quickly and picks up the pace as she describes the zoo of organisms in soils, from nematodes to the vast array of microbes. She emphasizes how bacteria and other microorganisms provide services that humans take for granted: filtering water, stabilizing soil, improving air quality and recycling nutrients that enable crops to grow. "I like to think of it as this factory underground," she says.

Wall credits her mother, a biology teacher, with helping to spark a lifelong interest in biology. Raised in Lexington, Kentucky, Wall

"MANY TIMES, I FELT LIKE I WAS HITTING THE GLASS CEILING AND GOT DISCOURAGED."

got her PhD in plant pathology from the University of Kentucky in her home town. In 1972, she left for the US west coast to pursue postgraduate research in nematology — convinced that nematode parasites had a lot to reveal about how life behaves above ground.

California was a shock at first. "That was eye-opening to me, because I had never crossed the Mississippi River, and it was — oh my god, where are the trees?" she says. But she ended up liking it there, and the University of California, Riverside, remained her home for much of the next two decades.

She strung together a series of grants to keep her work going, confident that soil microorganisms were more significant than most researchers realized. "Originally, I was just convinced these all make a difference and I was waiting to be proven wrong," she says.

Wall focused at first on nematodes in deserts and arid croplands, conducting the bulk of her research in southern California, New Mexico and Michigan. By the late 1980s,

she was seeking ways to understand a species' impact on an ecosystem. "If you want to find out how a plant parasite has an effect on a root or a predator, how do you exclude everything in the soil except that?"

She tried chemicals to kill off species, but they also harmed what she wanted to study. Then a colleague suggested that Wall go somewhere without plants, where the food web was simpler. "I tossed around a number of places," she says. "And we ended up in Antarctica."

She and her colleague Ross Virginia from Dartmouth College in Hanover, New Hampshire, decided to collect samples in the McMurdo Dry Valleys, a series of ice-free basins near the US McMurdo research station. The valleys receive no snow or rain, and humidity is so low that researchers have found the mummified remains of seals that made their way into the valleys thousands of years ago. Previous researchers had discovered nematodes and other life near glacier-fed streams that trickle during summer, but experts thought that the dry soils making up most of the valleys were barren.

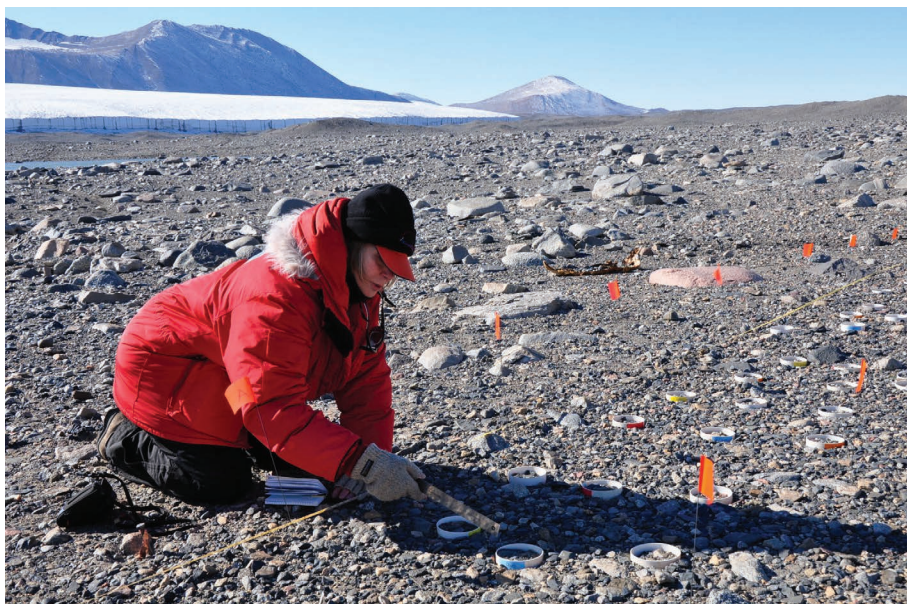
On one of Wall and Virginia's first visits to the Dry Valleys, they had just six hours to collect as many samples as possible before the helicopter returned to pick them up. They found nematodes in about 65% of the samples. "I couldn't believe it," she says. Ultimately, this showed that life can thrive even in the most inhospitable underground environments, revealing that major ecosystems were being overlooked.

Wall has returned to the Dry Valleys every field season except 1992, when she didn't receive funding for the trip. To recognize that long-running research, the US Geological Survey named valleys there after Wall and Virginia.

Their work in Antarctica dovetailed with discoveries that Wall had previously made about how nematodes cope with extremely dry conditions in the US Southwest. In the Chihuahuan Desert, Wall and her colleagues showed that the worms rely on anhydrobiosis¹: they shed most of their water and put metabolic activity on hold. Wall says that the nematodes end up looking like Cheerios, the ring-shaped dry cereal.

When she went to Antarctica, Wall and her colleagues found that Dry Valley nematodes use the same mechanism² to cope with arid conditions there³.

With one eye focused on tiny nematodes, Wall kept the other on the bigger picture of how these creatures fit into ecosystems. This was all part of her ever-growing desire to understand and highlight the importance of life underground — something routinely ignored by many researchers until roughly the past decade. Studies that tracked the decomposition of fallen leaves and other organic materials, for example, tended to



Wall at work in the Antarctic Dry Valleys.

overlook the role of soil organisms.

Wall says she grew tired of that limited perspective. “We wanted to show that animals are important in these processes.”

So in 2001, she started a global, multiyear project to measure the impact of soil animals. Her team sent mesh bags filled with hay to colleagues at more than 30 sites around the world. Placed in various locations, the bags attracted worms, beetles and other types of soil invertebrate, while control bags excluded them. Wall’s team then analysed the carbon content in each bag and compared the rates at which the organic matter decomposed with and without the soil animals. The results supported Wall’s point: soil fauna increased decomposition rates significantly in many regions⁴. A follow-up study⁵ found that excluding soil fauna reduced decomposition rates by a global average of 35%.

Those studies helped to convince researchers to pay more attention to life in soil. “We now understand how key these organisms are to many ecosystem processes,” says Amy Austin, an ecologist at the University of Buenos Aires.

The litter finding means that there could be big changes in how carbon moves throughout ecosystems as forces such as climate change alter soil communities.

Wall and her colleagues have seen some of this up close during their most recent field season in Antarctica. In as-yet-unpublished work, they found that the dominant nematode in the Dry Valleys, an endemic genus named *Scottinema*, has been declining in number, whereas a nematode that lives in wetter soils, *Eudorylaimus*, has been increasing, thanks to the melting of ice and permafrost. “It looks like there’s going to be a species shift,” she says. “It’s a fight for habitat.”

Scottinema is Wall’s favourite nematode. “It’s

living in this harshest environment, mostly by itself, and it’s just so recognizable,” she says. But that’s not the only reason that she has concerns about the species’ decline.

The two nematodes feed on different carbon sources in the soil, and population changes could alter the rate at which underground carbon escapes into the atmosphere. If so, the carbon-storage potential of the soil in Antarctica — a crucial region for absorbing carbon dioxide from the atmosphere — could change. Shifts in soil biota elsewhere on the planet could also affect how much carbon remains locked up, she says.

EMISSARY FOR SOIL

In August this year, Wall found herself at the White House talking about soils with other experts and policymakers as part of a national effort to prevent erosion and promote soil health. It was the latest scene in a role she has increasingly embraced over the past 15 years — to bring soil health to the global stage.

As Wall’s research career blossomed, she took on more leadership positions. She served as president of the American Institute of Biological Sciences in 1993 and the Ecological Society of America in 1999. By that point, her involvement in these organizations was making her think bigger. “I’d been pretty concentrated on the Antarctic research,” she says. “I thought I should be doing more.”

She began participating in and leading initiatives that were increasingly global in scope — chairing, for example, the International Biodiversity Observation Year starting in 2001, which funded research projects to highlight the importance of biodiversity around the world. In 2011, Wall became the founding science chair of the Global Soil Biodiversity Initiative, the group behind the soil atlas that was launched

in May. Looking forward, Wall wants to integrate data on soil health and biodiversity into global policies for mitigating large-scale environmental challenges. And she’s talking to colleagues about launching a big US experiment to unravel the relationships between soils, biodiversity and health.

“Conservation and protecting species is a very old idea, and so is soil conservation. But only now are these two ideas coming together,” she says.

While campaigning for soils, Wall has also been a champion for women in science. When she was starting out, there weren’t many role models for women in her field. And when she made her first trips to Antarctica, she made do with men’s long underwear and boots, and endured eight-hour flights on military aircraft that lacked sit-down toilets.

Wall was initially turned down for a tenure-track position at the University of California, Riverside, in the late 1980s — a decision that she and others suggest was related to her gender. Jill Baron, director of the North American Nitrogen Center at Colorado State University, says that how Wall recovered from that rejection is emblematic of her character. “She moved on into this stellar career,” says Baron. “And she’s been working to make sure that other young women who come in don’t have to ever have that again.”

That kind of drive makes a big impression on people just entering science. Ashley Shaw, a PhD student studying under Wall, recalls their first meeting. “She was just so enthusiastic about her science and what she was working on,” says Shaw. “I walked away feeling like I could save the world.”

Wall joined Colorado State University in 1993 to become director of the institution’s Natural Resource Ecology Laboratory. There, colleagues say, she attracted interdisciplinary, accomplished scientists, which elevated the stature of the lab both on and off campus. She now serves as founding director of the university’s School of Global Environmental Sustainability.

There, in an office covered in photos and paraphernalia from Antarctica, she talks eagerly about her goals for the future — and takes offence when people ask her if she plans to retire. “Whether I pass my physical to go to Antarctica or get too old and have to have a wheelchair dropped for me from the sky,” she says, “I want to keep working on the issues.” ■

Rachel Cernansky is a journalist in Denver, Colorado.

1. Whitford, W. G. *et al. Soil Biol. Biochem.* **13**, 417–425 (1981).
2. Treonis, A. M., Wall, D. H. & Virginia, R. A. *Funct. Ecol.* **14**, 460–467 (2000).
3. Wall, D. H. & Virginia, R. A. *Appl. Soil Ecol.* **13**, 137–150 (1999).
4. Wall, D. H. *et al. Glob. Chang. Biol.* **14**, 2661–2677 (2008).
5. García-Palacios, P., Maestre, F. T., Kattge, J. & Wall, D. H. *Ecol. Lett.* **16**, 1045–1053 (2013).

COMMENT



TECHNOLOGY The geeks and entrepreneurs of commercial space travel **p.304**

ECOLOGY Trees, a lesson in standing tall by standing together **p.306**

CHINA Sponge cities to retain run-off, control flooding, and reuse storm water **p.307**

OBITUARY Seymour Papert, father of educational computing **p.308**

DESIGN PICS INC/NGC



A wheat test plot in Maryland.

Agricultural R&D is on the move

Big shifts in where research and development in food and agriculture is carried out will shape future global food production, write **Philip G. Pardey** and colleagues.

The geographical distribution of food and agricultural research and development (AgR&D) is changing. Our analysis of more than 50 years of data indicates that the governments of middle-income nations are investing more than those of high-income ones for the first time in modern history. The numbers also suggest that, globally, private-sector spending on AgR&D is catching up with public-sector spending. Meanwhile, the gap between spending by high-income and low-income countries is widening.

Investments in R&D are inextricably intertwined with growth in agricultural productivity and food supplies¹. But it takes decades², not months or years, for the consequences of these investments to be fully realized. Today's R&D investment decisions will cast shadows forward to 2050 and beyond, making the trends we report here especially significant for the future of food production.

DATA GATHERING

To track shifts in where AgR&D occurs worldwide, we revised and updated the various data series on spending maintained by the University of Minnesota's International Science and Technology Practice and Policy (InSTePP) Center in St Paul. Successive versions of these series have been developed over decades by collating and harmonizing data obtained from many government and international agencies, private firms and unpublished sources, and using statistical approaches developed to infer missing observations³. Our global update took 6 years, and involved direct input from more than 60 collaborators at national and international statistical and scientific agencies.

Extensive details on the construction of our data series are available online (see go.nature.com/2cc9t4b). In short, the data include new and revised estimates of the amount of AgR&D spending by universities and government agencies for 158 countries from 1960 to 2011. They also include new global estimates of the amount of such R&D spending by private firms for three decades, from 1980 to 2011. (All spending in local currency units was converted to international dollars using 2009 purchasing power parity (PPP) exchange rates.)

These data reveal that we are in ►

FEEDING THE WORLD

More than 50 years of data show major shifts in who is spending what on food and agricultural research and development (AgR&D). For all countries, spending in local currencies is converted to dollars using 2009 purchasing power parity (PPP) exchange rates.

Low income Middle income High income

SHIFTING RANKS

For the first time in modern history, middle-income countries are investing more in public-sector AgR&D than are high-income ones.



INDUSTRY-LED

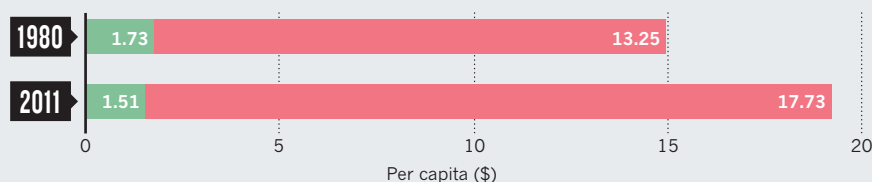
In both high- and middle-income countries, the share of AgR&D by private companies is increasing relative to that pursued by universities and government agencies.

Public AgR&D Private AgR&D



WORLDS APART

The gap between poor and rich countries in per capita spending on public AgR&D widened from 7.7-fold in 1980 to 11.7-fold in 2011.



► the midst of a historic transition (see 'Feeding the world'). For 2011, 5% of world-wide investment in all forms of R&D⁴ was directed towards food and agriculture. These global gross domestic public and private expenditures on AgR&D totalled US\$69.3 billion (in PPP dollars). Around 55% of this spending took place in high-income countries (as classified by the World Bank), down from 69% in 1980. Meanwhile, middle-income countries (including China, Brazil and India) were responsible for 43% (their share was 29% in 1980).

The shift in expenditures by universities and government agencies has been even more dramatic. Rich countries accounted for 56% of global public-sector spending on AgR&D in 1960, but only 47% in 2011. By this point, government spending in middle-income countries — 50% of global AgR&D public-sector spending — had overtaken that in high-income countries.

What is driving this shake-up in the rank order of spenders?

It is complex. Decades of decline in the real price of food and a sense that food provision was a solved problem may have fostered complacency among policy-makers and politicians in those countries that had a leading role in AgR&D throughout most of the twentieth century⁴ — including the United States, the United Kingdom and Australia. Meanwhile, some middle-income countries have been ramping up their spending to feed their increasingly wealthy populations (in the case of China and India), or to push into export markets (Brazil).

In middle-income nations overall, public spending grew by nearly 6% per year between 2000 and 2011, compared with an average of nearly 4% per year during the previous four decades. In rich countries, public AgR&D spending grew by just 0.8% between 2000 and 2011 (all figures are adjusted for inflation).

INDUSTRY INVESTMENT

Another major recent shift revealed by our data is in the balance between public- and private-sector contributions.

Historically, the bulk of research in food and agriculture was carried out by universities and government agencies. But in 2011, an average of 52.5% of the research on crop breeding, informatics, fertilizers, pesticides and food technologies in rich countries was being done by private firms (in 1980, the figure was 42%). For middle-income countries, the private proportion of domestic spending was 37% in 2011 (up from 19% in 1980). Middle-income countries' share of private AgR&D spending in 2011 was 35.5%, up from close to 16% in 1980.

The recent growth in investment in private AgR&D in China is especially striking. In 2011, more than \$6 billion, or

SOURCE: ORIGINAL ANALYSIS BY P.G.P. ET AL.

around 57% of the country's entire domestic AgR&D spending that year, came from the private sector. In China, the 'industrial enterprises' engaged in AgR&D include state-owned organizations — such as the China National Agricultural Development Corporation and the China National Cereals, Oils and Foodstuffs Corporation, which now effectively operate as for-profit firms — as well as private companies such as the Yili Group in Hohhot, and the China Yurun Food Group in Nanjing.

The increasing importance of private-sector R&D globally reflects two reinforcing developments. One is the impressive growth in R&D in crop genetics, farm machinery, agricultural chemicals and food processing in at least some middle-income countries. The other is the offshoring of AgR&D to rapidly growing middle-income countries by multinational firms headquartered in the rich countries. In recent years, firms that have opened R&D facilities in China have included Nestlé (which has three locations), Syngenta, PepsiCo and General Mills.

Today, the influence of corporations on AgR&D is vast. People tend to think of private-sector R&D — for instance, that pursued by companies such as Monsanto, DuPont Pioneer and Syngenta — as being mainly focused on agricultural chemicals, crop breeding and machinery. Yet food and beverages — involving multinationals such as PepsiCo, Kraft Heinz and Nestlé — was the focus of 44% of the total rich-country private AgR&D in 2011.

As countries become wealthier, people tend to eat out more and eat more processed and prepared foods⁵, and so returns on private-sector investments in food research are likely to be higher. Similar market opportunities open up for farm technologies such as improved seeds, fertilizers, herbicides and machinery as countries develop and farms typically consolidate and improve their physical access to urban markets.

THE R&D DIVIDE

Although the positioning of the top investors in AgR&D is shifting, little seems to have changed for those at the bottom. In fact, on a per capita basis, investment by low-income countries has shrunk considerably.

In 1980, for every dollar of AgR&D spent in high-income countries, just 3.5 cents was spent in the low-income countries. Three decades on, this divide is roughly the same. But the gap has widened considerably when AgR&D spending is evaluated per capita. In 1980, the rich countries invested \$13.25 per person in public AgR&D, whereas the poor countries invested \$1.73 (a 7.7-fold difference). By 2011, this per capita spending gap had widened to an 11.7-fold difference: rich countries invested \$17.73 per person, poor countries invested just \$1.51.

The divide is even more pronounced for private-sector spending. In 2011, for every dollar of private AgR&D spent in high-income countries, a meagre 0.8¢ was spent in low-income countries. Moreover, whereas private firms in rich countries spent \$1.10 for every public AgR&D dollar in 2011, the comparable private investment in poor countries was 15¢.

In short, those regions of the world that are experiencing the highest rates of population growth — the number of people living in sub-Saharan Africa has more than doubled since 1980 to 962 million today — are the places where per capita investment in AgR&D is among the lowest in the world.

A CHANGING WORLD

One of the major global challenges in the years ahead is getting the relevant agricultural innovations into the hands of the world's poor farmers, such as those in south Asia and sub-Saharan Africa. Even with the rise of some middle-income countries, food and agricultural research continues to be concentrated in just a handful of nations. In 2011, the top 10 countries ranked by spending on AgR&D accounted for 70% of the total investment worldwide; the bottom 100 contributed just 9% of that year's total. Yet these 100 are home to 22% of the world's population.

More and sustained government funding will be essential, along with robust and agile institutional innovations that foster public and private investment in poor-country agriculture. Without efforts to improve the global spread and adaptation of locally relevant technologies, it is likely to get much harder for poor farmers to feed themselves, let alone their nations' increasingly urbanized populations.

In those countries currently responsible for most of the world's agricultural production, the innovation challenges are also pressing, if different. History has already shown the cost of running down investment on food and agricultural research in the face of ever-evolving pathogens. The emergence of new virulent strains of wheat stem rust in Uganda in the late 1990s and their subsequent spread throughout Kenya, Ethiopia, South Africa and elsewhere in Africa is a reminder of the need for continued scientific vigilance⁶. Years of success in keeping the disease at bay had left only a handful of researchers worldwide studying the pathogen.

Without sufficiently supported research and innovation in agriculture, crop yields

are bound to decline as economic and environmental changes (including changes in weather patterns and crop pests and diseases driven in part by climate change) undermine past productivity gains.

Achieving even higher levels of productivity to feed a growing, increasingly wealthy and more urbanized population — while sustaining or rehabilitating fragile natural resources — is going to require considerably more investment in AgR&D. It will also require both public and private investment, because the two tend to support different, often complementary, types of R&D. The private sector is more attuned to market opportunities — and so well-suited to supply pesticides to farmers, for example. The public sector is better placed to investigate solutions to landscape-scale, longer-term challenges, such as the management of pesticide resistance.

If present trends continue, global AgR&D in the middle of the twenty-first century will look very different from how it looked at the dawn of the century. The rise of AgR&D in the rapidly growing middle-income countries, and the increase in private-sector participation in various regions are encouraging. But the retreat from public AgR&D by rich countries and the continued comparatively low levels of investment in many poorer countries, are concerning. Rapidly regaining lost ground for these parts of the world is an obvious priority if we are to feed the world sustainably to 2050 and beyond. ■

Philip G. Pardey is a professor in applied economics and director, and **Connie Chan-Kang** is a research associate at the

International Science and Technology Practice and Policy Center, University of Minnesota, St Paul, Minnesota, USA.

Steven P. Dehmer is research investigator and health economist at the *HealthPartners Institute, Minneapolis, Minnesota, USA.*

Jason M. Beddow died suddenly in April 2016 as this Comment was being finalized for submission.

e-mail: ppardey@umn.edu

1. Pardey, P. G., Chan-Kang, C., Beddow, J. M. & Dehmer, S. M. *Documentation — InStePP International Innovation Accounts: Research and Development Spending, version 3.5 (Food and Agricultural R&D Ser.)* (InStePP Center, 2016); available at <http://go.nature.com/2cc9t4b>
2. Alston, J. et al. *Persistence Pays: U.S. Agricultural Productivity Growth and the Benefits from Public R&D Spending* (Springer, 2010).
3. Hurley, T., Rao, X. & Pardey, P. G. *Am. J. Agric. Econ.* **96**, 1492–1504 (2014).
4. Pardey, P. G., Chan-Kang, C., Beddow, J. M. & Dehmer, S. P. *Am. J. Agric. Econ.* **97**, 1312–1323 (2015).
5. Fabiosa, J. F. in *The Oxford Handbook of the Economics of Food Consumption and Policy* (eds Lusk, J., Roosen, J. & Shogren, J.) Ch. 23 (Oxford Univ. Press, 2012).
6. Pardey, P. G. et al. *Science* **340**, 147–148 (2013).

Supplementary information accompanies this article online: see go.nature.com/2cfvkij

TECHNOLOGY

Selfies in space

Alexandra Witze examines a book on the techheads behind commercial high-altitude travel.



On 4 October 2004, US test pilot Brian Binnie flew a winged wonder to space and back. The rocket-powered *SpaceShipOne* won a US\$10-million prize for becoming the first commercial spaceplane — a breakthrough meant to herald the birth of the space-tourism industry. A decade later, its successor *SpaceShipTwo* broke apart during a test flight. Co-pilot Michael Alsbury had made a control error, and was killed as a result. The nascent business of take-a-selfie-in-space seemed as far from reality as ever.

Now, in *How to Make a Spaceship*, journalist Julian Guthrie tackles the story of private spaceflight. Readers who want to know about its early days will revel in her charismatic sketches of the space geeks, entrepreneurs and aviation buffs who made *SpaceShipOne*. Those looking for in-depth analysis of how that history relates to today's commercial spaceflight should look elsewhere. Space tourism remains a bucket-list thrill for billionaires, and NASA has adopted private spaceflight only to ferry science equipment, drinking water and extra bin bags up to the International Space Station.

Guthrie's earlier *The Billionaire and the Mechanic* (Grove Press, 2013) dealt with businessman Larry Ellison's obsessive quest to win the America's Cup yachting race. In *How to Make a Spaceship*, she deploys her skill to observe rich, driven enthusiasts racing to get to the edge of space. Despite the

stomach-wrenching microgravity and harsh radiation of space, humans have been pushing to go there since Soviet cosmonaut Yuri Gagarin became the first person to reach orbit in 1961.

First among the motley racers is Peter Diamandis, a serial dreamer who had helped to found a student space-exploration group and an international space university by the time he was 28. He is the kind of technofanatic who tracks how many days he has been alive and is entranced by the fact that the Massachusetts Institute of Technology numbers its buildings instead of naming them. Much of the book revolves around his efforts to jump-start a private space industry. Frustrated by cancelled space missions and hoping to open the final frontier to more than a select cadre of astronauts, Diamandis and a few like-minded techheads gathered in a Colorado mountain cabin in winter 1994. There they hatched the concept of a lucrative award to spur the development of commercial spaceflight. It became known as the

X Prize (see *Nature* **482**, 469; 2012), and Guthrie tracks Diamandis's attempts to find someone to fund it.

In 1996, he got an early boost from backers in St Louis, Missouri — the city whose investors had helped Charles Lindbergh to fly from New York to Paris on the first solo transatlantic flight some 70 years before. Diamandis went on to cobble together for the \$10-million lure. Enter the next group of dreamers — those looking to win the money. Of these, the most technically adept was Burt Rutan, the contrarian aviation designer who recorded many firsts, including building the broad-winged *Voyager* plane that circled the world in 1986 without stopping or refueling. Rutan runs an eclectic company in California's Mojave desert, where engineers and test pilots push the boundaries of flight. Backed by the deep pockets of Microsoft's Paul Allen, Rutan decided to make a bid for the X Prize. He dreamt up the innovative design in which *SpaceShipOne* dropped from a carrier plane, ignited its engines and coasted to the arbitrarily designated point where space begins, 100 kilometres up.

Guthrie's anecdotes illuminate Rutan's environment more than the enigmatic man himself. We glimpse the angst of the spouse left behind as a test pilot soars into the sky in an experimental machine. We hear of the engineer who in February 2003 listened aghast to the news of the space shuttle



How to Make a Spaceship: A Band of Renegades, and Epic Race, and the Birth of Private Spaceflight
JULIAN GUTHRIE
Penguin: 2016.

MARK GREENBERG/VIRGIN GALACTIC

Putative spaceplane *Columbia* coming apart, underscoring *SpaceShipTwo*, held under carrier plane that lives were at risk. *WhiteKnightTwo*.

Other colourful characters enter the narrative. They include John Carmack, the video-game designer who founded Armadillo Aerospace in Mesquite, Texas, to shoot for the prize; Steve Bennett, who sent his Starchaser rocket soaring above north-west England; and the Romanian Dumitru Popescu, who as an engineering student recruited his wife to help build rockets in his father-in-law's backyard. More famous names also emerge. Erik Lindbergh recreated his grandfather's flight in a modern plane to cope with the emotional pressure of his family's intense legacy and raise money for the prize. Entrepreneurs Anousheh and Amir Ansari used their personal fortune to sponsor the X Prize (and, eventually, to buy Anousheh a ride to the International Space Station in 2006). British billionaire Richard Branson ensured that an enormous Virgin logo was painted on the side of *SpaceShipOne* so that the television cameras would catch it in the morning light.

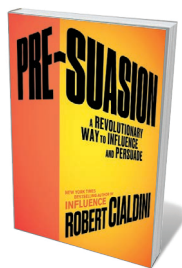
Guthrie sketches the interplay between these personalities as they jostle towards the X Prize deadline of December 2004. In the end, no competitor came close to the scrappy Rutan, who won the purse with two flights five days apart.

What remains unanswered is whether all this geekiness more than a decade ago has truly transformed commercial space-flight. In lieu of contributing to the X Prize, entrepreneur Elon Musk founded SpaceX in Hawthorne, California, which is now ferrying cargo to the space station (and will soon do the same with astronauts, along with aerospace company Boeing). Its competitor Blue Origin, set up by Amazon founder Jeff Bezos, was barely known in 2004 but has since pioneered reusable suborbital rockets that could save costs. There may or may not be a long-term business case for private spaceflight, but at the moment space tourism does not seem to be it.

Ultimately, *How To Make A Spaceship* is about the entrepreneurial work needed to launch such a project; short shrift is given to technical details and subsequent history. The fatal *SpaceShipTwo* accident is relegated to an epilogue, and an engine-test explosion that killed three of Rutan's employees in 2007 is not even mentioned. Yet Branson's Virgin Galactic continues to sell seats on future space flights for a quarter of a million dollars each. It expects to send a re-build of *SpaceShipTwo* into the skies on its first test flights later this year. ■

Alexandra Witze is a correspondent for Nature based in Boulder, Colorado.
e-mail: witzescience@gmail.com

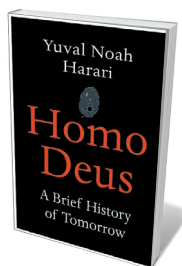
Books in brief



Pre-Suasion: A Revolutionary Way to Influence and Persuade

Robert Cialdini SIMON AND SCHUSTER (2016)

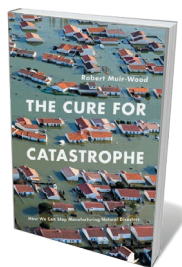
Fittingly, *Influence* (William Morrow, 1984) became one of the most influential studies in behavioural science, a triumph of field research on persuasion and how to resist it by social psychologist Robert Cialdini. Here Cialdini turns the tables, analysing how to harness persuasion by “frontloading” attention and pinpointing patterns of association conducive to change. His trove of findings and case studies covers how our focal points determine who we see as influential, how babies can be “pre-suaded” to be helpful, and how language can become a fulcrum in fraught negotiations.



Homo Deus: A Brief History of Tomorrow

Yuval Noah Harari HARVILL SECKER (2016)

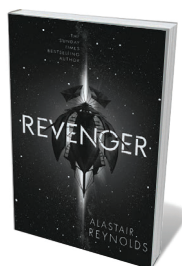
Historian Yuval Noah Harari's blockbuster *Sapiens* (Harvill Secker, 2014; see *Nature* **512**, 369; 2014) was a trenchant treatise on what he sees as our species' resistible rise to global dominion. In this equally acerbic forecast, Harari argues that the biological paradigm that casts organisms as biochemical algorithms shaped by natural selection could open the way to domination by networked computer algorithms. He opines that, as search engines and social media absorb our life histories and artificial intelligence advances, “dataism” may even make humanity obsolete.



The Cure for Catastrophe: How We Can Stop Manufacturing Natural Disasters

Robert Muir-Wood ONEWORLD (2016)

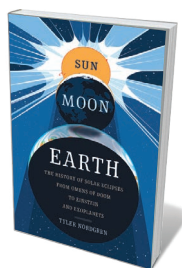
From the August earthquake in central Italy to the Fukushima crisis of 2011, multitudes of ‘natural’ disasters are exacerbated by shoddy construction, non-existent preparedness and political inertia. Disaster expert Robert Muir-Wood's study is science in the round, spanning centuries of catastrophes, key figures such as seismologist Charles Richter, forecasting, the intricacies of insurance (multistorey concrete buildings are revealed as “weapons of mass destruction” in a quake) — and a detailed, workable recipe for resilience.



Revenger

Alastair Reynolds GOLLANCZ (2016)

This latest science-fiction gem by astrophysicist Alastair Reynolds is a pacy space opera set in a far-future universe, where a broken civilization hangs on in a phalanx of artificial worlds. Rebellious teenagers Fura and Adrana join the crew of a solar-sailed vessel, riding the photon winds in search of lost technologies in the galactic deeps. Reynolds makes the human story compelling in a narrative that, spiced with bizarre characters aplenty and propelled by vengeance, smacks intriguingly of everything from Robert Louis Stevenson's *Treasure Island* to *Mad Max*.



Sun Moon Earth

Tyler Nordgren BASIC (2016)

On 21 August 2017, the United States will experience its first total solar eclipse in 40 years. Astronomer Tyler Nordgren's primer maps essentials for that event, contextualized by a fascinating history that sweeps us from Anaxagoras' explanation of eclipses in the fifth century BC to Arthur Eddington's test of Einstein's theory of general relativity during the May 1919 total eclipse. Nordgren is a wonderful guide to both the science and the sensory thrills, such as the shimmer of Bailey's beads or the eerie twilight of totality. *Barbara Kiser*



Trees communicate with each other using chemical signals carried on the breeze.

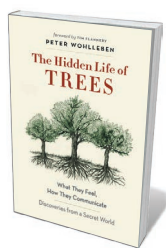
DENDROLOGY

The community of trees

Richard Fortey ponders a study that casts forests as exquisitely complex, multistorey networks.

The Ents — the tree beings in J.R.R. Tolkien's *The Lord of the Rings* — conducted leisurely conversations that ignored mere human timescales. Years might pass to allow their deeper ruminations. Tolkien understood that arboreal time plays out over centuries. The Ents also had moral values and took sides, and were capable of exercising will, forging alliances and showing affection. After reading *The Hidden Life of Trees*, I suspect that German forester Peter Wohlleben regards beeches, his favourite trees, as not unlike Ents.

Much new science has been woven into this engaging natural history. Trees are networkers. Far from the solitary splendour of the ancient old stager, it turns out that trees communicate with one another through their roots. Underground fungi — mycorrhizae associated with the root network — form a sort of subterranean internet that connects trees, passing messages and even nourishment between neighbours. Nor do trees passively tolerate the onslaught of insects on their tasty young leaves. Chemical



The Hidden Life of Trees: What They Feel, How They Communicate — Discoveries from a Secret World
PETER WOHLLEBEN
Greystone: 2016

signals carried on the breeze from infested trees cause forest fellows to crank up their own chemical armouries. It's not a case of every tree for itself: the forest can behave as a single entity when it yields a great crop of acorns or beechnuts, or lies fallow for a year. Trees share a common response to weather and nourishment. The hidden network allows for the nurturing of small trees in the understorey, where too little light penetrates for effective photosynthesis. The 'children' of a 'parent' tree bide their time until the oldest topples and the understorey underdogs at last get the chance to reach for the skies. Wohlleben's capable description of the leisurely drama of the forest through

generations of trees idealizes an ecological progression too often interrupted by human felling and woodland management. It seems that trees are both more cooperative within a species and more complicated within a lifetime than prejudice might allow.

Wohlleben's vision of life among the trees has been developed during his decades-long stewardship of a chunk of forest dominated by beech in the Eifel, a mountain range straddling Germany and Belgium. He clearly desires woodlands to return to a state in which the slow life cycles of the trees are allowed to run without interference — a regrowth of the European 'wildwood' that grew up as the climate recovered after the retreat of the last Ice Age. He presents this as a golden age of arboreal life. Trees age at their own pace and die, to be replaced by 'family' that has been sheltering in their shadows for many years, nurtured on the mycelial teat. It's a kind of utopia for Ents. Not one such undisturbed ancient forest survives in Europe, except possibly in the Białowieża Forest of Poland. In Britain, woods have been managed since the Iron Age.

Whatever the virtues of this scenario, I have problems with Wohlleben's narrative approach. He describes trees as if they possessed consciousness. During times of drought they make "cries of thirst" or "might be screaming out a dire warning to their colleagues". They experience "rising panic". A seedling's growth is portrayed as fratricide as it sees off its siblings. It is rather extraordinary to read a book centred on co-evolution without a mention of natural selection. After a while, the urge to attribute motivation to the behaviour of trees becomes irksome. It is not so far away from hugging trees to connect to a supposed deeper reality.

Wohlleben sets out his stall quite specifically: "The distinction between plant and animal is, after all, arbitrary." Well, no, it's not. It has been a fact of phylogenetic separation for more than 1.7 billion years, during which exceedingly long time the two kingdoms have followed their own paths. Yes, problems in common require comparable solutions: communication and nutrition are universals, as scholars such as the plant biochemist Anthony Trewavas have shown. It is of selective advantage to trees to share news of insect threats, just as antelope respond together to the twitch of a lion's tail.

Trees are splendid and interesting enough in their own right without being saddled with a panoply of emotions. The anthropomorphism in this otherwise compelling book is more spice than it needs. Trees ain't Ents. ■

Trees are splendid and interesting enough in their own right without being saddled with a panoply of emotions. The anthropomorphism in this otherwise compelling book is more spice than it needs. Trees ain't Ents. ■

Richard Fortey is a research associate at the Natural History Museum, London. His latest book is *The Wood for the Trees, an essay on the history and natural history of a small English beechwood*. e-mail: r.fortey@nhm.ac.uk

Correspondence

Harness passion of private fossil owners

Reproducing palaeontological results depends on unrestricted access to fossils described in the literature, allowing others to re-examine or reinterpret them. Museums have policies and protocols for keeping materials in the public trust, but accessibility to privately owned fossil collections can be a problem.

For example, the existence of an important early bat fossil in a private collection was long known, but it was only after a second specimen was acquired and made available by a museum that researchers published a description of it (N. B. Simmons *et al.* *Nature* **451**, 818–821; 2008). Another example is the unique fossil of a supposed four-legged 'snake', also privately owned, that was made temporarily available through a private German museum and then withdrawn after its description was published (D. M. Martill *et al.* *Science* **349**, 416–419; 2015).

We suggest that the enthusiasm of private collectors for their valuable and spectacular fossils should instead be harnessed by researchers, to the benefit of both parties. For example, scientists can invite collectors to participate in their projects and be co-authors on the publications (R. R. Reisz *et al.* *Sci. Nat.* **102**, 50; 2015), or they can name the new species after the collector (S. P. Modesto *et al.* *Proc. R. Soc. B* **282**, 20141912; 2015) — all on the condition that the specimen is donated to an institution with public right of access.

Robert R. Reisz *University of Toronto Mississauga, Canada.*
Michael W. Caldwell *University of Alberta, Edmonton, Canada.*
robert.reisz@utoronto.ca

Species can be named from photos

As an international group of taxonomists who study a range of taxa, we consider that you

misconstrued the case of a new insect species that was described on the basis of photographs (see *Nature* **535**, 323–324; 2016).

The species was described without a preserved type specimen, the individuals having escaped before preservation (S. A. Marshall and N. L. Evenhuis *ZooKeys* **525**, 117–127; 2015). The *International Code of Zoological Nomenclature* allows for this — the authors (included here as signatories) followed the letter and the spirit of the *Code*, giving a description and a formal species name. It was based on material that supported their conclusions and an explanation of the circumstances to justify naming a species without an extant type. Peer reviewers judged the data sufficiently reliable to anchor a species name.

As you point out, a physical specimen has features that might not be captured in a photo. However, types are name-bearers, not “standards for species delimitation” (D. S. Amorim *et al.* *Zootaxa* **4137**, 121–128; 2016). Significant knowledge about a species may build up before we can properly preserve a name-bearing type. The *Code* allows for the naming of those species.

More than 90% of the planetary biota still awaits description. We need to adopt new technologies while recognizing that museum specimens and nomenclatural stability are crucial for taxonomy.

Thomas Pape *Natural History Museum of Denmark, Copenhagen.*
tpape@snm.ku.dk

**Supported by 34 signatories (see go.nature.com/2cur7a6 for full list).*

China's sponge cities to soak up rainwater

China's Sponge City programme aims to improve resilience to urban expansion and climate change by enabling cities to save and resupply rainwater. It is crucial for cities such as Beijing and Jinan, which suffer water shortages even after severe flooding. However, several

hurdles must be overcome to get it working efficiently.

The programme will involve some 30 pilot cities this year (see www.mohurd.gov.cn). They will create a 'sponge' infrastructure to detain runoff, control flooding, recharge groundwater and reuse storm water. The project still has to recruit enough planners, designers and construction workers to support this colossal initiative. Time is short for completing technical training.

Plans and technology will need to be customized for individual cities, where local weather conditions and the degree of urbanization can vary considerably; a blanket strategy will not work.

Once in place, the sponge infrastructure should be combined with conventional drainage systems, particularly in areas of medium- and high-intensity urbanization.

Dasheng Liu *Ecological Society of Shandong, Jinan; and Ludong University, Yantai, China.*
liu_sdiep@126.com

Clearing the way for reef destruction

Agricultural practices are accelerating the health decline of Australia's Great Barrier Reef, affecting marine and terrestrial ecosystems (see also S. L. Maxwell *et al.* *Nature* **536**, 143–145; 2016). Last month, intensive opposition from the agricultural lobby blocked new legislation by the Queensland government that would have protected the reef's catchment areas from land clearing — despite support for the legislation from almost 500 scientists (go.nature.com/2cnlftg).

Broad-scale land clearing tripled after the state relaxed its vegetation regulations in 2013 (see go.nature.com/2cjm6zm). Subsequent assurances by the state that it would reduce land clearing contributed to last year's decision by the United Nations Educational, Scientific and

Cultural Organization not to add the Great Barrier Reef to its 'World Heritage In Danger' list.

The new regulations would have protected the pristine woodlands of Cape York and reduced terrestrial runoff, promoting recovery of those parts of the Great Barrier Reef that have been severely affected by unprecedented coral bleaching.

Preserving what remains of the reef's world-renowned biodiversity depends on urgently forging effective agreements with Queensland's agricultural sector.
April E. Reside *University of Queensland, Brisbane, Australia.*
Tom C. L. Bridge, Jodie L. Rummer *James Cook University, Townsville, Queensland, Australia.*
a.reside@uq.edu.au

Avoid bias against junior researchers

I disagree with Joy Burroughs-Boenisch's proposal that journal reviewers and editors, as well as English-language editors, should be informed when papers are to be assessed as part of a higher degree (*Nature* **536**, 274; 2016).

The (student) status of an author is irrelevant to whether the science is of sufficient quality to justify publication. A declaration of student status could entrench bias against junior scientists who already have few, if any, publications on which to build a reputation.

I also question whether service providers who assist in the publication process warrant listing in a PhD thesis statement. Editors, for example, improve the quality of the science through appropriate peer review and — along with copy editors and English-language editors for translated texts — optimize its presentation through clarification and technical correction. However, they are not part of the scientific advance that justifies publishing the paper in the first place.

Andrew K. Skidmore *University of Twente, the Netherlands.*
a.k.skidmore@utwente.nl

Seymour Papert

(1928–2016)

Father of educational computing.

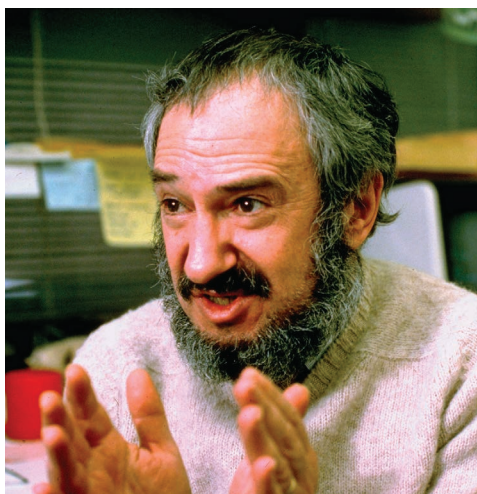
In the mid-1960s, when few people had even seen a computer, Seymour Papert was making it possible for children to use and program them. He spent his career inventing the tools, toys, software and projects that popularized the view of computers as incubators of knowledge.

Papert wrote three seminal books on using the computer to supercharge learning, aimed at academics, teachers and parents — *Mindstorms: Computers, Children, and Powerful Ideas* (1980), *The Children's Machine: Rethinking School in the Age of the Computer* (1993) and *The Connected Family: Bridging the Digital Generation Gap* (1996). Few academics of Papert's stature have spent as much time as he did working in real schools. He delighted in the theories, ingenuity and playfulness of children. Tinkering or programming with them was the cause of many missed meetings.

Papert, who died on 31 July, was born on leap day 1928 in Pretoria, South Africa. His father was an entomologist. Before he was two, he became enamoured by automotive gears that were lying around his home, which became the basis for early maths and science experiences. As an educator, he sought to help each learner to find his or her 'gears': objects or experiences they could mess about with, intuiting powerful ideas along the way. Papert believed that what gears could not do for all, the computer, the Proteus of machines, might.

Papert was repelled by apartheid. He ran afoul of the authorities by organizing classes for local black servants while in school. His anti-apartheid activities as a young adult branded him a dissident and prohibited him from travelling outside South Africa. He earned a bachelor's degree in philosophy (1949) and a PhD (1952) in mathematics at the University of the Witwatersrand in Johannesburg.

Without a passport, in 1954 he made his way to the University of Cambridge, UK, where he earned a second doctorate, in 1959, for work on the lattices of logic and topology. From 1959 to 1963, Papert worked at the University of Geneva with the Swiss philosopher and psychologist Jean Piaget. Their collaboration led to great insights into how children learn to think mathematically. Papert built on Piaget's theory of constructivism with a learning theory of his own: constructionism. It proposed that the best way to ensure that knowledge is built in the learner is through the active construction of something



shareable — a poem, program, model or idea.

In 1963, artificial-intelligence (AI) pioneer Marvin Minsky invited Papert to join him at the Massachusetts Institute of Technology (MIT) in Cambridge. Papert was soon promoted to co-direct Minsky's Artificial Intelligence Laboratory. The pair co-authored the 1969 book *Perceptrons*; their mathematical analyses of how neuron-like networks comprised of individual agents could model the brain had a great impact on AI research. In 1985, Papert became a founding faculty member of the MIT Media Laboratory, where he led research groups on epistemology and learning and the future of learning.

Thinking about thinking and the freedom to achieve one's potential were the leitmotifs of his life. He wanted to create "a mathematics children can love rather than inventing tricks to teach them a mathematics they hate".

In the late 1960s, Papert was among the creators of Logo, the first programming language for children. One element that made Logo accessible was the turtle, which acted as the programmer's avatar. As mathematical instructions were given to the turtle to move about in space, the creature dragged a pen to draw a trail. Such drawings created turtle geometry, a context in which linear measurement, arithmetic, integers, angle measure, motion and foundational concepts from algebra, geometry and even calculus were made concrete and understandable. Mathematics became playful, personal, expressive, relevant and purposeful.

In 1968, Alan Kay, now known as the designer of what became the Macintosh graphical user interface, was so impressed by

the mathematics that he saw children spontaneously engaged in at Papert's Logo lab at MIT that on his flight home, he sketched the Dynabook, the prototype for what became the personal computer. In 1989, Australian schools seeking to realize Papert's ideas began providing a laptop to every student. In 2000, Maine governor Angus King proposed providing a laptop for every 7th and 8th grader (typically 12–14-year-olds). Papert spent two years making the case across the state, causing popular opinion to override legislative resistance. The programme remains in place today. Papert was also an inspiration behind the One Laptop per Child initiative that has reached millions of children in the developing world.

A 1971 paper co-authored by Papert, 'Twenty Things to Do with a Computer' (see go.nature.com/2buuwe), marks the birth of the modern 'maker movement'. It describes a world in which children would create by programming inventions and experiments outside of a PC. In the mid-1980s, Papert and his colleagues made that world a reality with the first programmable robotics system for children, LEGO TC Logo. The name of LEGO's current line of robotics sets — Mindstorms — is a hat-tip to Papert. In 1989, LEGO endowed a permanent chair at the MIT Media Lab in his name.

Although critical of institutional schooling, Papert's research took place in schools, often with under-served populations of students. In 1986, he was invited to help Costa Rica reinvent its educational system, and from 1999 to 2002, Papert led an alternative, high-tech, project-based learning environment inside a prison for teenagers.

Papert dared educators to grow, invent and lead in a system prone to compliance and standardization. He argued that education is a natural process that blossoms in the absence of coercion.

In Papert's eyes, the computer was an object to think with. He built a bridge between progressive educational traditions and the Internet age to maintain the viability of schooling, and to ensure the democratization of powerful ideas. ■

Gary S. Stager is the co-author of *Invent To Learn — Making, Tinkering, and Engineering in the Classroom*, and curator of the online Papert archive, dailypapert.com. He worked with Seymour Papert from 1986 to 2006. e-mail: gary@stager.org

BILL PEECE/LIFE/GETTY

EARTH SCIENCE

Extraordinary world

The isotopic compositions of objects that formed early in the evolution of the Solar System have been found to be similar to Earth's composition — overturning notions of our planet's chemical distinctiveness. [SEE LETTERS P.394 & P.399](#)

JAMES M. D. DAY

Classes of meteorites called chondrites formed from material in the solar nebula — the cloud of gas and dust from which the planets also formed — and have not subsequently undergone substantial mineralogical or geological changes¹. Earth's bulk composition, including its metallic core and silicate mantle and crust, should be the same as that of the bulk Solar System, and would therefore be expected to correspond to chondrite-meteorite compositions. This idea was cast into doubt in 2005 by the spectacular finding² that terrestrial materials have elevated abundances of neodymium-142 (¹⁴²Nd) compared with ordinary or carbonaceous chondrite meteorites, implying that Earth's accessible regions cannot have a chondritic composition. Burkhardt *et al.*³ (page 394) and Bouvier and Boyet⁴ (page 399) now reassess this matter by reporting high-precision analyses of isotopes of neodymium and samarium (Sm) in a variety of materials from the early Solar System. Their findings reaffirm that Earth's composition is chondritic.

Nature has gifted us with a spectacular diversity of stable isotopes and of isotopes formed through radioactive decay, thus allowing us to place precise constraints on Earth's composition. For example, the ratio of the

abundance of ¹⁴²Nd to that of ¹⁴⁴Nd partly reflects the decay of ¹⁴⁶Sm to ¹⁴²Nd; ¹⁴⁶Sm is a now-extinct isotope that had a half-life of 103 million years⁵. Today's accessible terrestrial mantle has a ¹⁴²Nd/¹⁴⁴Nd ratio greater than that of ordinary and carbonaceous chondrites, which are thought to be the building blocks of the planets in the Solar System. A 'missing reservoir' of terrestrial material has been invoked to explain the difference. This reservoir has a lower ¹⁴²Nd/¹⁴⁴Nd ratio than have chondrites, and must have been isolated from the accessible Earth within the first 20 million to 30 million years of terrestrial formation².

Various models have been proposed to explain the location of the missing reservoir — for example, perhaps it is isolated in the deepest parts of Earth², or formed an early crust that was lost to space during massive planetary collisions⁶. The existence of such a reservoir would have profound implications for terrestrial evolution and habitability, because the reservoir would also contain a large fraction of the radioactive, heat-producing elements (uranium, thorium and potassium) and therefore would fundamentally affect Earth's heat budget⁷.

An alternative possibility is that the discrepancy between Earth and chondrites was caused by isotopic variations imparted during the formation of the Solar System. Our Solar System

is composed of elements inherited from extinct stars and supernovae that seeded different proportions of isotopes into different parts of the early solar nebula, leading to isotopic compositions in chondrites that are distinct from that of the present-day Earth. For example, ¹⁴²Nd is formed in stars by the slow process (s-process), a particular type of nucleosynthetic process in which neutrons are captured by atomic nuclei at a relatively pedestrian pace. The other isotopes of neodymium are formed either by the s-process or in 'core-collapse' supernovae by rapid neutron capture (the r-process).

Samarium isotopes also show composition variations associated with the s- and r-processes, and so, by combining analyses of samarium and neodymium isotopes, geochemists have a powerful tool for understanding the isotopic compositions inherited from stars and supernovae. By measuring samarium and neodymium isotopes at high precision in chondrites, Burkhardt *et al.* show that the materials that formed Earth were enriched in neodymium formed by the s-process.

Taking a different approach, Bouvier and Boyet measured the isotopic composition of calcium–aluminium inclusions — the earliest objects to have condensed from the solar nebula and therefore the oldest solid objects in the Solar System. They find that plots of the isotopic compositions of these inclusions,

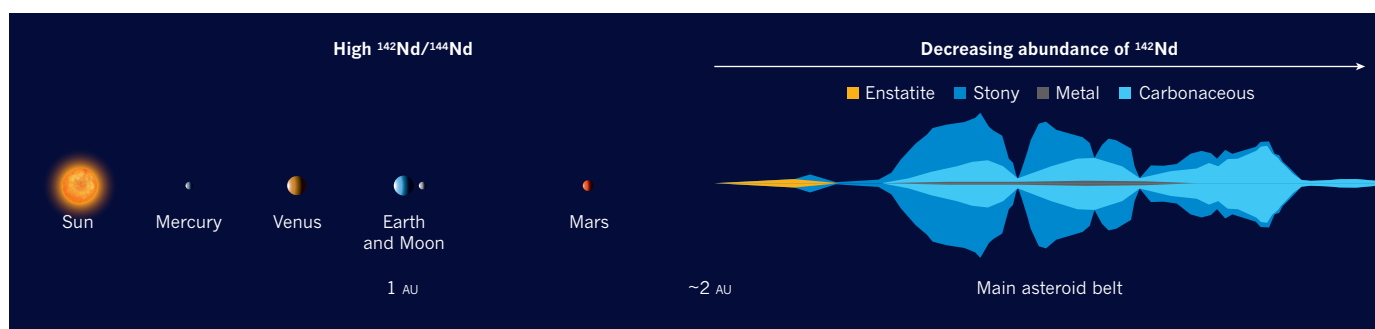


Figure 1 | An isotopically heterogeneous Solar System. Evidence that Earth has a higher abundance of neodymium-142 (¹⁴²Nd) than do two classes of meteorite (ordinary and carbonaceous chondrites) suggests that the Solar System inherited an uneven distribution of isotopes from stars and supernovae during its formation. Burkhardt *et al.*³ and Bouvier and Boyet⁴ measured ¹⁴²Nd abundances in early Solar System objects whose compositions broadly correlate with those of three types of main-belt asteroid that have surfaces similar to enstatite chondrites, stony chondrites or carbonaceous chondrites. Taking

into account the distribution of asteroids and planets, the results imply that the abundance of ¹⁴²Nd decreases with distance from the Sun. Like Earth and the Moon, Mars has a high ¹⁴²Nd/¹⁴⁴Nd ratio compared with that of chondrites⁸. Distances from the Sun are in astronomical units (1 AU is 150 million kilometres) and are shown to scale, but planetary-body sizes are not to scale. The distribution of asteroids in the main belt is shown as an exaggerated cross-section; the distribution of metal asteroids, which have surfaces similar to those of iron meteorites, is shown for completeness. (Adapted from ref. 11.)

and of some enstatite chondrites (a rare form of meteorite), pass straight through the isotopic composition of modern Earth's accessible mantle. This implies that nucleosynthetic variations explain the distinctions between some chondrite groups and Earth. Remarkably, using two distinct approaches, both studies come to the same conclusion: after corrections for nucleosynthetic effects, Earth must have chondritic abundances of samarium and neodymium, and, by association, also of uranium, thorium, potassium and some other elements.

Even so, Earth remains extraordinary. Enstatite chondrites and calcium–aluminium inclusions cannot have formed Earth; they might partly overlap isotopically with our planet, but they are either too depleted in key elements, including volatile elements, or too chemically reduced to explain the terrestrial composition. Conversely, carbonaceous and ordinary chondrites are too depleted in ^{142}Nd (ref. 3). Moreover, chondrites and their components are unlikely to be representative of the materials that formed Earth because the materials that fed the formation of the inner planets no longer exist, having been processed within Mercury, Venus, Earth, the Moon or Mars. Instead, the formation of the Solar System might have been the result of 'cosmic cookery', in which stars and supernovae sprinkled variable isotopic compositions into the mix, akin to a chef seasoning a dish before serving it.

Burkhardt *et al.* and Bouvier and Boyet propose that the relative abundance of ^{142}Nd decreases with distance from the Sun (Fig. 1), and that this reflects either variable processing of dust from the solar nebula or distinct compositions of material sprinkled from star and supernova sources into the solar nebula over time. Further high-precision measurements of Martian meteorites⁸ could help to validate these proposals: if the authors are correct, then such meteorites would be slightly deficient in ^{142}Nd compared with Earth. If Mercurian or Venusian meteorites can be found, then they would be expected to be enriched in ^{142}Nd compared with Earth, allowing stringent constraints to be defined for an isotopically heterogeneous, and possibly radially stratified, Solar System.

The latest results have major implications for our understanding of Earth's evolution. Earlier theories of a modern terrestrial mantle with a high $^{142}\text{Nd}/^{144}\text{Nd}$ ratio — dubbed a supra-chondritic Earth — have led to models that interpret lavas with primitive noble-gas and lead isotope compositions as coming from one of the most ancient accessible mantle reservoirs⁹. If Earth actually has a chondritic Sm/Nd ratio^{3,4}, then the missing reservoir hypothesized in previous studies^{2,6,7,9} never existed. Alternatively, the 'primitive' isotopic variations measured in some lavas reflect the complex assimilation of ancient crustal materials into magmas formed by partial melting of Earth's mantle¹⁰. ■

James M. D. Day is at the Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0244, USA. e-mail: jmdday@ucsd.edu

- Anders, E. & Grevesse, N. *Geochim. Cosmochim. Acta* **53**, 197–214 (1989).
- Boyet, M. & Carlson, R. W. *Science* **309**, 576–581 (2005).
- Burkhardt, C. *et al.* *Nature* **537**, 394–398 (2016).
- Bouvier, A. & Boyet, M. *Nature* **537**, 399–402 (2016).

- Meissner, F., Schmidt-Ott, W.-D. & Ziegeler, L. Z. *Phys. A* **327**, 171–174 (1987).
- Campbell, I. H. & O'Neill, H. St C. *Nature* **483**, 553–558 (2012).
- Jellinek, A. M. & Jackson, M. G. *Nature Geosci.* **8**, 587–593 (2015).
- Caro, G., Bourdon, B., Halliday, A. N. & Quitté, G. *Nature* **452**, 336–339 (2008).
- Jackson, M. G. *et al.* *Nature* **466**, 853–856 (2010).
- Day, J. M. D. *Chem. Geol.* **440**, 91–100 (2016).
- DeMeo, F. E. & Carry, B. *Nature* **505**, 629–634 (2014).

EVOLUTION

Teenage tetrapods

Bone analysis of aquatic tetrapods from around the time when these four-limbed vertebrates began to move onto land reveals that the large specimens were only juveniles, raising questions about how these animals developed. SEE LETTER P.408

NADIA B. FRÖBISCH

One of the most fascinating topics in vertebrate evolution is the transition of finned fish to four-limbed tetrapods. This transition involved changes to many aspects of the biology of our fish ancestors, including their respiration, waste removal and skeletal system¹. The evolution of the limb, which eventually led to our own arms and legs, was a prerequisite for tetrapods' conquest of the land and ability to evolve the amazing variety of body forms and means of locomotion observed in both extinct and modern vertebrates. Given the pivotal role of this move onto land, the anatomical transformations involved

have been a major focus of research, not only in palaeontological studies, but also in studies of evolutionary developmental biology and the relationship between anatomical structures and their function^{2,3}. A paper on page 408 by Sanchez *et al.*⁴ reveals insights into growth patterns of the early tetrapod *Acanthostega*. The results will provide a deeper understanding of the development and evolution of our four-legged forerunners.

Although many advances have been made in understanding the evolutionary transition from fish to tetrapod, a key piece of the puzzle has remained elusive — how did the earliest tetrapods grow? The process by which an organism develops from the fertilized egg to

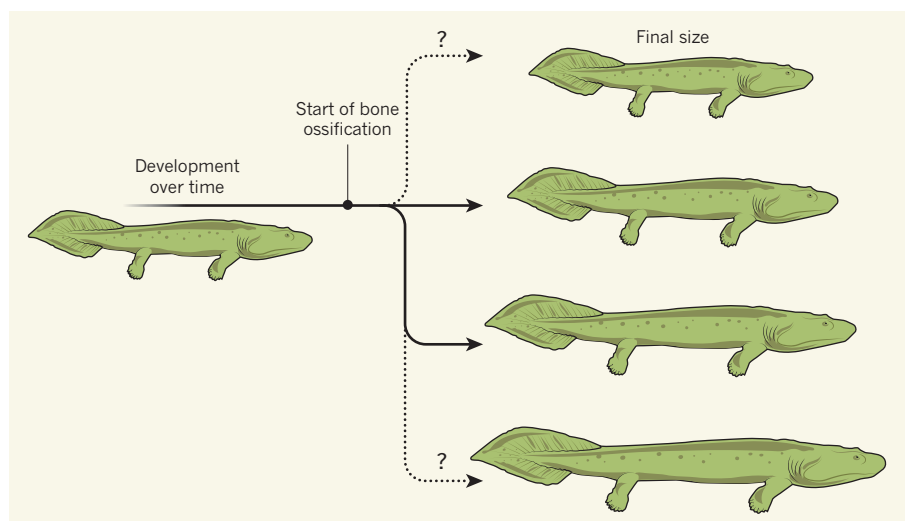


Figure 1 | Possible developmental pathways in the tetrapod *Acanthostega*. Sanchez *et al.*⁴ analysed the internal microstructure of fossil forelimb bones of juvenile *Acanthostega*. The authors found that ossification began at a late stage, after the animals had grown to nearly full size, and that there were at least two size classes (solid arrows). The two classes could represent differences in body size for male and female forms, or developmental plasticity in response to intrinsic or environmental factors. Further size classes might have existed (dotted arrows), which could potentially be revealed by larger sample sizes.

the adult form (known as ontogeny) reveals details about the evolution and biology of a species that cannot be made by studying adult individuals alone. Series of fossils that chart the development of later tetrapods from larvae to adults⁵ have provided a wealth of developmental data. However, the ontogenetic development of the earliest tetrapods has been poorly understood because such information is rare in the fossil record, and juvenile and adolescent stages had not been identified.

Sanchez and colleagues' study animal, *Acanthostega*, is one of the earliest known tetrapods, and lived about 365 million years ago during the Devonian period. The authors used synchrotron microtomography, a non-destructive way to generate 3D structural representations of the microstructure of fossil bone, and studied the long upper bone of *Acanthostega*'s forelimb, the humerus. Bone is a dynamic tissue, and studying its microstructure can reveal unique information about the physiology, growth and life history of vertebrates, because the internal structures provide indications of how fast an animal grew, how old an individual was and when growth ceased. Sanchez *et al.* investigated *Acanthostega* samples from a fossil assemblage site in which the individuals had all died together, probably in a drought following a catastrophic flood event. Although only a few humeri were available, they provide a glimpse into the growth patterns of this transitional species.

In tetrapods, the humerus initially forms as a cartilage precursor, with bone material being subsequently deposited in a process known as ossification. Surprisingly, Sanchez and colleagues' imaging data indicate that all of the specimens they investigated were still in the juvenile growth phase and had not reached sexual maturity. Even more surprisingly, *Acanthostega* seemingly reached almost its final size while retaining a cartilaginous humerus during an early-juvenile period that lasted several years (near final size was inferred when the bone microstructure showed that growth had slowed substantially). By contrast, ossification of the limb bones in modern tetrapods starts much earlier than in either *Acanthostega* or our fish predecessors. The finding that *Acanthostega* grew to almost final size and still had a cartilaginous humerus supports the hypothesis that the earliest tetrapods had a predominantly, if not an exclusively, aquatic lifestyle, because a cartilaginous humerus would probably have been unable to bear much weight. This indicates that limbs initially served a purpose on land other than locomotion.

However, the most compelling of Sanchez and colleagues' results lies in a clear disjunction between size and degree of ossification — some individuals reached the same degree of ossification in the long bones at a much smaller body size than others. Developmental plasticity, an organism's capacity to respond flexibly to different external cues

throughout life, is thought to have an important role in evolution^{6,7}. Studies of fossils and modern amphibians have elucidated the complex and fascinating connections between developmental plasticity and the responses of individuals to cues of population dynamics and environmental factors, including competition between juveniles, length of growth period, climatic factors and predation^{8–11}. Sanchez *et al.* identified two size classes in their study (Fig. 1), although the small sample size limits interpretations with respect to possible drivers of plasticity. It is possible that there were more size classes, which may be revealed when further samples are available.

In *Acanthostega*, the decoupling of size and degree of ossification in a long juvenile stage could indicate that developmental plasticity, and possibly alternative life-history strategies, were already present in the earliest tetrapods. A high degree of developmental plasticity might have provided the means for our early ancestors to respond to changing intrinsic and environmental conditions, and could thereby have had a central role in the initial evolutionary success and subsequent diversification of tetrapods. ■

Nadia B. Fröbisch is at the Museum für Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, 10115 Berlin, Germany.

e-mail: nadia.froebisch@mf-n-berlin.de

1. Clack, J. A. *Gaining Ground: The Origin and Evolution of Tetrapods* 2nd edn (Indiana Univ. Press, 2012).
2. Schneider, I. & Shubin, N. H. *Trends Genet.* **29**, 419–426 (2013).
3. Standen, E. M., Du, T. Y. & Larsson, H. C. E. *Nature* **513**, 54–58 (2014).
4. Sanchez, S., Tafforeau, P., Clack, J. A. & Ahlberg, P. E. *Nature* **537**, 408–411 (2016).
5. Fröbisch, N. B., Olori, J. C., Schoch, R. R. & Witzmann, F. *Semin. Cell Dev. Biol.* **21**, 424–431 (2010).
6. Moczek, A. P. *et al. Proc. R. Soc. B* **278**, 2705–2713 (2011).
7. West-Eberhard, M. J. *Developmental Plasticity and Evolution* (Oxford Univ. Press, 2003).
8. Schoch, R. R. *Annu. Rev. Earth Planet. Sci.* **37**, 135–162 (2009).
9. Schoch, R. R. *Evolution* **63**, 2738–2749 (2009).
10. Urban, M. C., Richardson, J. L. & Freidenfelds, N. A. *Evol. Appl.* **7**, 88–103 (2014).
11. Whiteman, H. H. *et al. Oecologia* **168**, 109–118 (2012).

This article was published online on 7 September 2016.

IMMUNOLOGY

Cytotoxic T cells that escape exhaustion

T cells of the immune system mount antiviral responses, but if a response fails, a chronic viral infection can develop. It now seems that a T-cell subset in lymphoid immune tissues can control chronic infection. SEE LETTERS P.412 & P.417

CINDY S. MA & STUART G. TANGYE

Although T cells are known to fight chronic viral infections, the exact requirements for this process have been a mystery. Papers in this issue by Im *et al.*¹ (page 417) and He *et al.*² (page 412) and in *Nature Immunology* by Leong *et al.*³ have identified a population of T cells that express the CXCR5 and CD8 surface proteins and may control these infections. These CXCR5⁺CD8⁺ T cells are located in immune tissues known as the secondary lymphoid system. The studies highlight the need to reassess the function of CD8⁺ T-cell subsets in controlling chronic viral infections.

Cytotoxic CD8⁺ T cells kill virus-infected cells and cancer cells⁴, which they target by recognizing specific molecules called antigens. It is thought that CD8⁺ T cells are absent from areas of secondary lymphoid tissues termed follicles, which are rich in immune B cells and are dedicated to generating antibodies. Follicles might therefore offer a location in which

viruses can evade T-cell attack and so create a viral reservoir that could sustain a chronic infection⁵. The expression of CXCR5 enables a different type of T cell, known as a T follicular helper cell, to migrate into an area in these follicles called the B-cell zone^{6–8}. T follicular helper cells also express the inhibitory receptor protein PD-1, and require a specific network of transcription factors for their development^{6–8}. Although CXCR5 had been detected on a small fraction of CD8⁺ T cells (less than 2%) in human blood and tonsils^{9,10}, the function of these cells was unclear. This has now been addressed by the current studies.

An experimental system for studying chronic viral infection is infection of mice with lymphocytic choriomeningitis virus (LCMV). LCMV infection causes a state known as immunological exhaustion, in which immunological attack on the infection is compromised because the CD8⁺ T cells that target virus-infected cells show reduced production of immune cytokine signalling molecules, together with reduced proliferation and

cytotoxicity¹¹. T-cell exhaustion occurs through various immune-cell regulatory pathways, including those that act through PD-1-mediated restraint of immune responses¹¹.

The authors of the three papers discovered that, during the chronic phase of LCMV infection, up to 30% of activated CD8⁺ T cells expressed CXCR5 (Fig. 1). Im *et al.* and He and colleagues found that these cells were located exclusively in lymphoid tissues, and were involved in controlling the infection. The CXCR5⁺CD8⁺ T cells were found to share many features^{6–8} and regulatory pathways^{8,12,13} with T follicular helper cells, such as high expression of co-stimulatory receptor proteins and transcription factors. By contrast, the three studies found that the expression of immune inhibitory receptors associated with exhausted CD8⁺ T cells¹¹ was reduced in CXCR5⁺CD8⁺ T cells compared with exhausted CD8⁺ T cells that did not express CXCR5 (CXCR5[−]CD8⁺ T cells). PD-1 expression was found to be reduced (by Im *et al.* and He *et al.*) or at a similar level (Leong *et al.*) on CXCR5⁺CD8⁺ T cells compared with CXCR5[−]CD8⁺ T cells.

These exhausted CXCR5[−]CD8⁺ T cells were found throughout the ‘red pulp’ and T-cell-rich regions known as T-cell zones in the spleens of infected mice. However, all three studies found that CXCR5⁺CD8⁺ T cells homed to T-cell zones, and Im and colleagues and Leong *et al.* also found them in B-cell zones. He *et al.* and Leong *et al.* identified CXCR5⁺CD8⁺ T cells in humans infected with HIV or Epstein–Barr virus, both of which can establish chronic viral infection. Chronic viral infection in both humans and mice yields distinct responses of CD8⁺ T cells depending on the cells’ CXCR5 status. Studies in mice clearly showed that CXCR5[−]CD8⁺ T cells acquired characteristics of exhaustion and were distributed across both immune and non-immune tissues, whereas CXCR5⁺CD8⁺ T cells were confined to lymphoid tissues and showed features that were intermediate between those of T cells that had not been activated by antigen and those of exhausted T cells.

Im and colleagues and He *et al.* found that CXCR5⁺CD8⁺ T cells were better at controlling infection than CXCR5[−]CD8⁺ T cells. Transfer of LCMV-targeting CXCR5⁺CD8⁺ T cells into virus-infected mice resulted in a substantially lower viral load than when CXCR5[−]CD8⁺ T cells were transferred. He *et al.* found that the numbers of CXCR5⁺CD8⁺ T cells in people with HIV were inversely correlated with the viral load in the blood. The results suggest that CXCR5⁺CD8⁺ T cells are predominantly responsible for controlling chronic viral infections in humans and mice.

Intriguingly, gene-expression profiling of mouse CXCR5⁺CD8⁺ T cells by Im *et al.* revealed a molecular signature resembling that of blood stem cells. Consistent with this, these authors found that the cells proliferated extensively and yielded CXCR5[−]CD8⁺

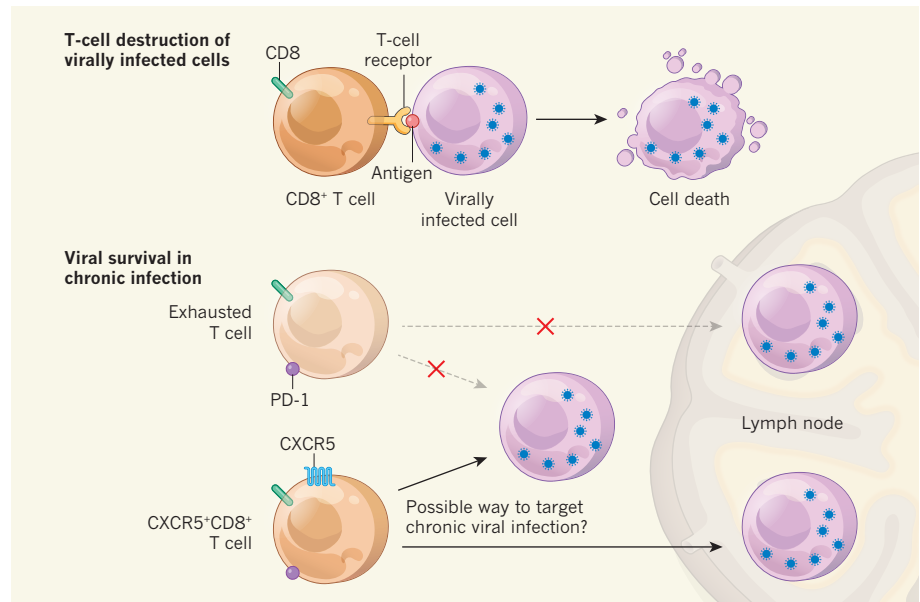


Figure 1 | A T-cell subset that targets chronic viral infection. When an immune T cell expressing CD8 protein (a CD8⁺ T cell) is exposed to a specific viral antigen molecule that it recognizes through the T-cell receptor, the T cell becomes activated and can kill the virally infected cell presenting the antigen. In chronic infection, infected cells can escape T-cell-mediated destruction if T cells enter an ‘exhausted’ state in which proteins such as PD-1 inhibit T-cell activation. Viruses could also escape destruction in locations where T cells are absent, such as proposed T-cell-free areas including the B-cell zone of immune follicles in lymph nodes. Studies by Im *et al.*¹, He *et al.*² and Leong *et al.*³ have identified a population of activated CD8⁺ T cells that express the receptor protein CXCR5 and target chronic viral infection. These CXCR5⁺CD8⁺ T cells are not fully exhausted. The cells might be located in regions previously thought to be T-cell free, although the studies give conflicting reports of the location of the T cells.

T cells, indicating that CXCR5⁺CD8⁺ T cells are both stem-cell-like cells and precursors of exhausted T cells.

Expression of PD-1 is a key determinant of whether T cells become exhausted. Leong and colleagues found that CXCR5⁺CD8⁺ and CXCR5[−]CD8⁺ T cells expressed comparable levels of PD-1, whereas Im *et al.* and He *et al.* found lower PD-1 levels in CXCR5⁺CD8⁺ T cells than in CXCR5[−]CD8⁺ T cells. Furthermore, Im *et al.* and He *et al.* found that the CXCR5⁺CD8⁺ T cells responded more robustly than CXCR5[−]CD8⁺ T cells to treatment with the antibody anti-PD-L1, which releases CD8⁺ T cells from the suppressive effects of PD-1. This ‘checkpoint inhibitor’ treatment can curb immune-system impairment, and is yielding stunning results as immunotherapy for various cancers¹⁴.

He *et al.* observed a synergistic antiviral effect when anti-PD-L1 treatment was combined with transfer of CXCR5⁺CD8⁺ T cells into infected mice. The effect was not seen with CXCR5[−]CD8⁺ T cells, suggesting that the antiviral effect of PD-1 inhibition is mediated exclusively by CXCR5⁺CD8⁺ T cells. Therefore, boosting the function of CXCR5⁺CD8⁺ T cells through PD-1 blockade is an attractive prospect for improving therapies for chronic viral infections, at least for infections that target lymphoid cells. If CXCR5⁺CD8⁺ T cells are responsible for viral eradication through an enhanced immunological response following PD-1 blockade, this raises questions about

the function of CXCR5[−]CD8⁺ T cells and why these exhausted cells are retained during responses to chronic viral infections.

Although the three studies agreed on many aspects of the characterization of CXCR5⁺CD8⁺ T cells, a major difference was where these cells are located in the secondary lymphoid tissues. Im *et al.* convincingly demonstrated that the cells reside in the T-cell zone, but not in B-cell follicles, and that CXCR5[−]CD8⁺ T cells occur mainly in the splenic red pulp. These authors propose that CXCR5[−]CD8⁺ T cells kill virus-infected cells in splenic red pulp and non-lymphoid tissues, whereas CXCR5⁺CD8⁺ T cells recognize virus-infected cells in T-cell zones of lymphoid tissue. By contrast, He and colleagues and Leong *et al.* found that mouse CXCR5⁺CD8⁺ T cells occur with B cells in the B-cell zone of lymphoid tissue, and Leong and colleagues even found them in close association with HIV-infected cells in B-cell follicles of human lymph nodes.

Unlike the other studies, Leong *et al.* found that CXCR5⁺CD8⁺ T cells expressed less cytotoxic molecules than CXCR5[−]CD8⁺ T cells, and there was no difference in viral loads between mice that received LCMV-specific CD8⁺ T cells from animals whose cells were able to express CXCR5 or from animals that lacked the gene enabling them to do so. Thus, in this experiment, whether or not the cells could express CXCR5 had no effect on the T cell’s antiviral function. These

findings predict less efficacy of CXCR5⁺CD8⁺ T cells in viral control, especially in infection with HIV and Epstein–Barr virus, which respectively persist in follicles in immune T cells that express the CD4 protein and in B cells. Im and colleagues' finding of an absence of CXCR5⁺CD8⁺ T cells in follicles and Leong and colleagues' finding of reduced cytotoxicity of CXCR5⁺CD8⁺ T cells compared with CXCR5⁺CD8⁺ T cells are consistent with the idea that B-cell follicles provide a 'sanctuary' for HIV-infected T follicular helper cells⁵. The precise function and location of CXCR5⁺CD8⁺ T cells remain unresolved.

These studies provide insights into the spatio-temporal and dynamic nature of CD8⁺ T-cell-mediated immunity against chronic infections. Parallels in the findings between humans and mice underscore the probable importance of CXCR5⁺CD8⁺ T cells in controlling protracted infections. Furthermore, the expansion and enhanced function of CXCR5⁺CD8⁺ T cells following PD-1 blockade render these cells a target for immune intervention when treating infectious disease. However, it is unknown whether these T cells are also generated in response to chronic viruses that infect non-lymphoid cells, such as hepatitis B or hepatitis C. It will be interesting to discover whether the cells infiltrate lymphoid and non-lymphoid tumours and so could also be targeted for cancer immunotherapy¹⁴. More studies of CXCR5⁺CD8⁺ T cells will be required before we can harness their full therapeutic potential. ■

Cindy S. Ma and Stuart G. Tangye are in the Immunology Division, Garvan Institute of Medical Research, Darlinghurst, New South Wales 2010, Australia, and at St Vincent's Clinical School, University of New South Wales. e-mails: c.ma@garvan.org.au; s.tangye@garvan.org.au

1. Im, S. J. *et al. Nature* **537**, 417–421 (2016).
2. He, R. *et al. Nature* **537**, 412–416 (2016).
3. Leong, Y. A. *et al. Nature Immunol.* <http://dx.doi.org/10.1038/ni.3543> (2016).
4. Cui, W. & Kaech, S. M. *Immunol. Rev.* **236**, 151–166 (2010).
5. Fukazawa, Y. *et al. Nature Med.* **21**, 132–139 (2015).
6. Tangye, S. G., Ma, C. S., Brink, R. & Deenick, E. K. *Nature Rev. Immunol.* **13**, 412–426 (2013).
7. Crotty, S. *Immunity* **41**, 529–542 (2014).
8. Crotty, S. *Annu. Rev. Immunol.* **29**, 621–663 (2011).
9. Forster, R., Emrich, T., Kremmer, E. & Lipp, M. *Blood* **84**, 830–840 (1994).
10. Quigley, M. F., Gonzalez, V. D., Granath, A., Andersson, J. & Sandberg, J. K. *Eur. J. Immunol.* **37**, 3352–3362 (2007).
11. Wherry, E. J. & Kurachi, M. *Nature Rev. Immunol.* **15**, 486–499 (2015).
12. Choi, Y. S. *et al. Nature Immunol.* **16**, 980–990 (2015).
13. Shaw, L. A. *et al. Nature Immunol.* **17**, 834–843 (2016).
14. Topalian, S. L., Drake, C. G. & Pardoll, D. M. *Cancer Cell* **27**, 450–461 (2015).

This article was published online on 24 August 2016.

GENOMICS

Geography matters for *Arabidopsis*

A free database describes genome sequences, gene expression and molecular modifications to DNA for more than 1,000 *Arabidopsis thaliana* plants, providing valuable information on the complex history and current variation of this species.

OUTI SAVOLAINEN & MARTIN LASCoux

The number of genomic resources is increasing rapidly, but large collections of high-quality, whole-genome sequences are available for only a few species, including humans¹ and fruit flies². Such collections can help researchers to address both basic and applied genetic questions. Writing in *Cell*, the 1001 Genomes Consortium³ and Kawakatsu *et al.*⁴ describe an advanced genomic resource for the thale cress (*Arabidopsis thaliana*), molecular biology's most prominent plant model.

In the first study, the consortium presents whole-genome sequences of more than 1,300 genetically different individuals (accessions) from a worldwide collection (Fig. 1). In plants, such data sets have so far been generated only for cultivated species, such as rice⁵ and tomatoes⁶. The *Arabidopsis* sequences are of high quality and easily accessible to users. The two papers jointly provide a good overview of all types of DNA-sequence variation.

Molecular modifications to DNA, such as the addition of methyl groups to the base

cytosine, can influence gene expression and thereby alter physical and biological traits. In the second paper, Kawakatsu *et al.* improve on earlier studies (for example, ref. 7) by recording such epigenetic variants in about 1,000 accessions, which largely overlap with the sequenced accession set from the first paper.

What sets this work apart from most other collections of whole-genome sequence data is that seeds of all these accessions are freely available to the scientific community. *A. thaliana* is a predominantly selfing species — offspring receive two identical copies of each gene from the parent, whereas in outcrossing species, the copies from the mother and father can be different. The genome sequence of each accession of a selfing species is therefore maintained in subsequent generations. As such, researchers can obtain seeds whose genomes have already been fully characterized. Furthermore, the accessions also have data on gene expression and methylation status in some environments⁴. The groups have produced a tremendously valuable resource.

A crucial issue for this kind of large-scale study is the distribution of samples over



Figure 1 | A catalogue of variation for *Arabidopsis thaliana*. Two studies^{3,4} document genome sequences, gene expression and molecular modifications to DNA for more than 1,000 varieties of this plant species.

JAMES MANN/ABRC

different geographical regions. The *Arabidopsis* accessions cover a large geographical area, but the sampling is much more uneven than that achieved for the project's human counterpart, the 1000 Genomes Project¹. It reflects past collecting efforts and largely focuses on the Iberian peninsula in southwest Europe and on Sweden.

In analyses of the genetic distance between plants, the consortium identified many Iberian accessions and one accession each from the Cape Verde Islands, Canary Islands, Sicily and Lebanon as relicts — representatives of populations that had given rise to plants that migrated to other regions but that had not migrated themselves. Most of the relicts from outside Iberia may in fact have arisen through dispersal by humans, because the collection sites historically had close cultural associations with the Iberian peninsula. The remaining non-relict samples are the result of the rapid, nearly worldwide expansion of the species, and enable studies of more-recent adaptation.

Although this analysis sheds new light on the movements of *A. thaliana*, more-extensive sampling is still needed to fully understand the history of the species. In particular, the source of the current widely distributed non-relict accessions needs further study. European populations of many other non-relict plant species (such as forest trees) are known to be derived from multiple colonizations from areas containing relicts (such as the Iberian, Italian and Balkan peninsulas)⁸.

Kawakatsu *et al.* combine the two groups' extensive data set to analyse how interactions between genetic and environmental factors contribute to changes in gene expression or methylation. For instance, the authors compare the extensive Swedish and Iberian populations and show that northern populations have more DNA methylation than southern ones — an interesting finding that should lead to more-detailed study. The consortium presents an analysis of genetic variants associated with variation in flowering time that identifies a few well-known sequence regions (loci) that regulate flowering at two different temperatures in the greenhouse.

Such overall analyses of large data sets have limitations as well as advantages, however. Previous studies (for example, ref. 9) showed that, in natural conditions, variation in flowering time is associated with a large set of loci that have small individual effects. The complex population history reflected in the current set of accessions provides an explanation for this discrepancy. Combining data from all over the world can result in false positives — perhaps, in correcting for this, the authors have eliminated many true-positive associations⁹. Furthermore, different loci may account for variation in different accessions, and such local associations would probably not be detected in an overall analysis. Clearly, regional studies, which involve less variability between accessions, are also needed, along with experimental work.

More generally, a combination of approaches will be required to fully understand the genetic basis of variation in adaptive traits¹⁰.

Researchers working with other plant species will benefit from the *Arabidopsis* resources. The effects of selection can be more easily tracked in closely related outcrossing species that have a simpler population history, such as *Capsella grandiflora*¹¹. But data from *A. thaliana* will help researchers to interpret their findings in *C. grandiflora*, and will provide opportunities for evolutionary comparisons. Even for more distantly related species, such as long-lived conifers (for which genomic data are only now becoming available^{12,13}), information about functional variation in *Arabidopsis* will be helpful.

Finally, although *Arabidopsis* is a selfer, most plant species are outcrossing. Research into the population genomics of these species will be aided by the analytical tools developed for species such as fruit flies and humans that have an outcrossing mating system¹². Genomic resources for individual species therefore provide wide-ranging benefits for researchers in many fields. ■

MATERIALS SCIENCE

Not so creepy under stress

Nanocrystalline alloys have excellent low-temperature mechanical strength but poor high-temperature resistance to creep — deformation due to continuous stress. An alloy has been made that overcomes this problem. SEE LETTER P.378

JONATHAN CORMIER

Metallic materials that have high-temperature applications in gas turbines require a combination of excellent mechanical strength and resistance to creep — deformation resulting from long-term applied stress. Nowadays, nickel-based 'superalloys' are key materials in the hottest sections of gas turbines, largely on account of their exceptional creep resistance at temperatures of up to 1,100 °C (almost 90% of their melting temperature)¹. The use of these superalloys has led to a spectacular decrease in the fuel consumption of gas turbines over the past 40 years, reducing their environmental impact². Searching for new lightweight metallic materials of even greater high-temperature capability than the superalloys is one of the biggest challenges the aerospace industry faces in reducing its contribution to global warming³. On page 378, Darling *et al.*⁴ report the development of a nanocrystalline alloy that combines impressive mechanical strength with

Outi Savolainen is in the Department of Genetics and Physiology, and in Biocenter Oulu, University of Oulu, 90014 Oulu, Finland. **Martin Lascoux** is in the Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, 752 36 Uppsala, Sweden. e-mail: outi.savolainen@oulu.fi

1. The 1000 Genomes Project Consortium. *Nature* **526**, 68–74 (2015).
2. Lack, J. B. *et al.* *Genetics* **199**, 1229–1241 (2015).
3. The 1001 Genomes Consortium. *Cell* **166**, 481–491 (2016).
4. Kawakatsu, T. *et al.* *Cell* **166**, 492–505 (2016).
5. 3000 Rice Genomes Project. *GigaScience* **3**, 7 (2014).
6. Lin, T. *et al.* *Nature Genet.* **46**, 1220–1226 (2015).
7. Dubin, M. J. *et al.* *eLife* **4**, e05255 (2015).
8. Petit, R. *et al.* *Science* **300**, 1563–1565 (2003).
9. Brachi, B. *et al.* *PLoS Genet.* **6**, e1000940 (2010).
10. Savolainen, O., Lascoux, M. & Merilä, J. *Nature Rev. Genet.* **14**, 807–820 (2013).
11. Williamson, R. J. *et al.* *PLoS Genet.* **10**, e1004622 (2014).
12. Nystedt, B. *et al.* *Nature* **497**, 579–584 (2013).
13. Neale, D. B. *et al.* *Genome Biol.* **15**, R59 (2014).

This article was published online on 7 September 2016.

high-temperature creep resistance.

Nanocrystalline metals and alloys are made of minuscule grains with diameters typically smaller than 100 nanometres⁵. Because of this structure, nanocrystalline materials have excellent mechanical strength at low temperatures (up to a few hundred degrees Celsius). However, the poor creep resistance of such materials has always prevented their use for high-temperature applications⁶.

Darling and colleagues' nanocrystalline alloy is based on a system consisting of copper grains with average diameters of about 50 nm (Fig. 1). The authors introduce particles of the metal tantalum with diameters of between 3 and 32 nm to the boundaries between the grains. To process the alloy, the copper and tantalum particles are first mechanically milled for 4 hours at a very low temperature (–196 °C) to produce a powder with the desired particle sizes. The powder is then pushed through a channel at 700 °C to form a bar of the alloy called a billet. The authors repeat this last process four times, which results in severe plastic deformation of the alloy — the

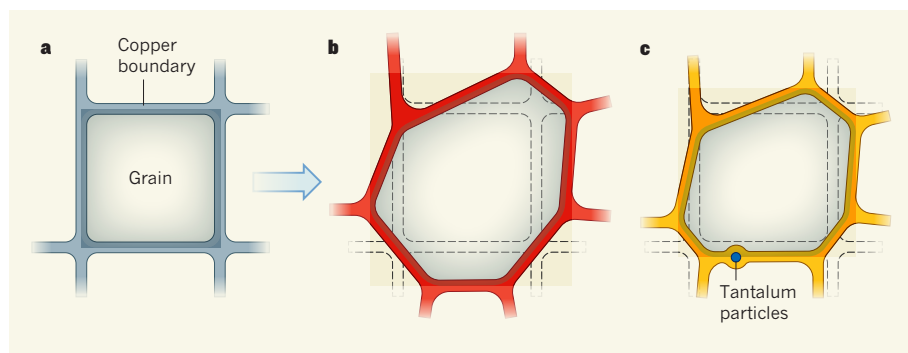


Figure 1 | A creep-resistant nanocrystalline alloy. Darling *et al.*⁴ construct a nanocrystalline alloy that has high-temperature resistance to creep — deformation under continuous stress. **a**, The alloy is based on a system of grains that are separated by boundaries made of copper atoms (a single grain is illustrated here). **b**, When long-term stress is applied to a typical system, the copper atoms diffuse to new positions, increasing the size of the grain. **c**, The authors add tantalum particles to a grain boundary. When long-term stress is applied, the size of the grain does not increase as significantly as in the case of pure copper — the alloy has greater creep resistance.

billet is 460% longer by the end of the process, ensuring a fine grain size.

The alloy's exceptional creep properties result mainly from the stability of its microstructure — the tantalum particles pin down the grain boundaries, preventing them from moving to new positions in response to stress at high temperatures. The authors find that the rate at which creep occurs in the alloy is six to eight orders of magnitude lower than in most other nanocrystalline metals, implying a spectacular improvement in durability. However, the alloy is not a direct candidate material for the hottest sections of gas turbines because it has high strength and creep resistance only at rather low temperatures (up to 600 °C), compared with the current nickel- and cobalt-based superalloys. But its development opens the door for new types of nanocrystalline alloy, provided that several key issues are addressed.

The first main concern for the industrial use of these nanocrystalline alloys is the processing route, especially in the aerospace industry, which requires reliable and stable processes. For alloys such as that developed by Darling *et al.*, a uniform dispersion of grain-boundary pinning particles, as well as a controlled grain size over a large volume, would be necessary for high-temperature components in gas turbines such as blades, vanes or disks. However, it would probably be difficult to achieve these two properties using the authors' processing route (particularly during the milling and severe plastic-deformation stages).

A second concern is resistance to oxidation, another design criterion for high-temperature industrial applications. Increasing the density of grain boundaries in metallic systems generally enhances oxidation because it raises the rate of grain-boundary diffusion⁷. However, a nanocrystalline alloy could be made more resistant to oxidation if a dense and protective outer oxide layer were to be rapidly grown. This could be achieved by adding to the alloy

elements such as aluminium or chromium, which are widely known to improve the environmental resistance of metallic materials⁷. Alternatively, a specific coating might be developed for the nanocrystalline alloy. Such a coating would need to be chemically compatible with the alloy (having similar chemical composition and thermal expansion) and contain high levels of aluminium and chromium to ensure excellent oxidation and corrosion resistance in harsh environments.

Finally, the authors' alloy would need to retain outstanding strength and creep resistance at temperatures considerably higher than 600 °C

to be used for the above-mentioned gas-turbine components. Instead of copper, the grains could be made of elements such as nickel or cobalt, which have a higher melting point and a stable crystallographic structure over the entire temperature range required. With the addition of a second source of strengthening (such as intragranular precipitation⁸, a process that provides efficient obstacles to the irreversible elongation of the grains), the creep properties of the alloy might reach the level of the nickel-based superalloys that are used today, but with greater mechanical strength. This increased strength could allow load-bearing sections of the components to be made smaller, and hence vastly lower in weight — improving gas-turbine efficiency and reducing fuel consumption. ■

Jonathan Cormier is at *École Nationale Supérieure de Mécanique et d'Aérotechnique, Institut Pprime, UPR CNRS 3346, 86961 Futuroscope-Chasseneuil, France.*
e-mail: jonathan.cormier@ensma.fr

1. Pollock, T. M. & Tin, S. J. *Propul. Power* **22**, 316–374 (2006).
2. Reed, R. C. *The Superalloys: Fundamentals and Applications* (Cambridge Univ. Press, 2006).
3. Pollock, T. M. *Nature Mater.* **15**, 809–815 (2016).
4. Darling, K. A. *et al. Nature* **537**, 378–381 (2016).
5. Kumar, K. S., Van Swyghoven, H. & Suresh, S. *Acta Mater.* **51**, 5743–5774 (2003).
6. Meyers, M. A., Mishra, A. & Benson, D. J. *Prog. Mater. Sci.* **51**, 427–556 (2006).
7. Birks, N., Meier, G. H. & Pettit, F. S. *Introduction to the High-Temperature Oxidation of Metals* 2nd edn (Cambridge Univ. Press, 2006).
8. Ardell, A. J. *Met. Trans. A* **16**, 2131–2165 (1985).

AGEING

Dietary protection for genes

Dietary restriction is known to extend lifespan in many species. It has now been shown to reduce DNA damage and extend lifespan in mice modelling human DNA-repair disorders. [SEE LETTER P.427](#)

JUNKO OSHIMA & GEORGE M. MARTIN

The accumulation of DNA damage is an inevitable side effect of living, and is one of the main causes of cellular and organismal ageing. Compromised DNA repair leads to persistent DNA damage, causing age-related disorders and shortening lifespans. In humans, this can manifest as progeroid syndromes, in which children or adults age at a greatly accelerated rate. On page 427, Vermeij *et al.*¹ demonstrate that a relatively modest degree of dietary restriction can greatly increase the lifespans of two mouse models of these human syndromes.

The authors' mouse strains harbour

mutations in genes involved in a DNA-repair process called nucleotide excision repair (NER). In one strain, a mutation reduces production of the protein ERCC1, which normally forms a complex with a DNA endonuclease enzyme to create DNA breaks and excise damaged sequences. ERCC1 mutations can cause three diseases in humans² — the accelerated-ageing disorders Cockayne syndrome and XFE progeroid syndrome, and xeroderma pigmentosum, in which people are extremely sensitive to DNA damage by sunlight. In the other mouse strain, a mutation inhibits production of another DNA endonuclease, called XPG. Human XPG mutations can present

as xeroderma pigmentosum and Cockayne syndrome³.

It has previously been shown⁴ that mice harbouring mutations in *Ercc1* exhibit many of the metabolic responses to stress that are seen in healthy mice subjected to dietary restriction — in both, biological pathways involved in physiological maintenance are enhanced at the expense of pathways involved in growth. This is thought to be a survival response that helps to protect NER-deficient mice. Vermeij *et al.* therefore investigated whether dietary restriction could enhance these protective responses in their animal models. Indeed, a 30% restriction led to a substantial increase in lifespan in both strains of mouse, as compared with siblings given unlimited access to food (those fed *ad libitum*).

A weakness of many investigations into dietary restriction is their failure to carefully investigate physiological and structural traits of the organism under study, which together can be used to gauge healthy lifespan. Vermeij and colleagues' study is a welcome exception, because it investigates a wide range of relevant traits — including those involving the brain and neuromuscular systems, which are particularly vulnerable to damage in human DNA-repair disorders⁵. A striking finding was that mutant mice subjected to dietary restriction retained 50% more neurons than did siblings fed *ad libitum*. Moreover, markers of DNA damage were reduced in the diet-restricted animals (Fig. 1), and transcriptional profiles were better preserved.

There were several other interesting findings. For instance, the authors showed that, in ERCC1-deficient mice fed *ad libitum*, genes that encode large proteins were more damaged than those that encode small ones. This makes sense, because DNA damage occurs randomly. As such, long genes suffer disproportionate amounts of stochastic damage. As another example, the weights of NER-deficient mice fed *ad libitum* gradually decreased over time, and Vermeij *et al.* found that these animals died when they reached around the same weight as diet-restricted mutants, which initially lost weight rapidly but then maintained a constant weight. Again, this makes sense — weight loss in mutants fed *ad libitum* reflects physiological decline, whereas initial weight loss related to scheduled dietary restriction actually enhances physiology.

Dietary restriction has long been known to extend healthy lifespan in many animal species⁵. In usual ageing, its effects are modulated mainly through inhibition of the IGF1 and mTOR molecular signalling pathways⁶, which have roles in nutrient sensing. IGF signalling is already suppressed in NER-deficient mice², so it comes as something of a surprise that the defects seen in these animals can be partially rescued by dietary restriction. Nonetheless, the authors confirmed that the IGF1 and mTOR pathways are further suppressed in the

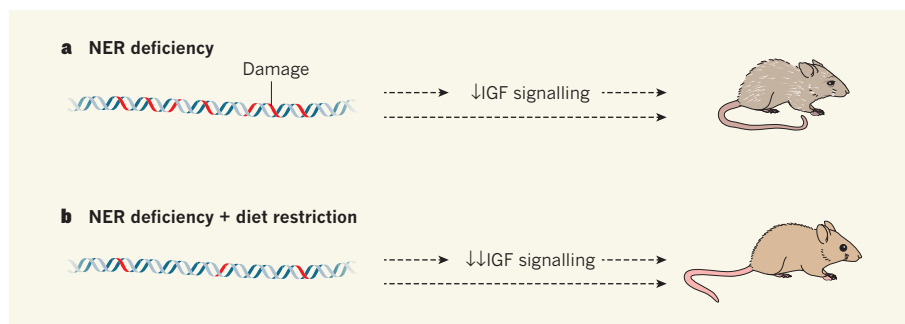


Figure 1 | Living better for longer. **a**, Nucleotide excision repair (NER) is a process by which DNA damage is repaired. Mice that harbour genetic mutations in the genes *Ercc1* or *Xpg* are NER deficient. DNA damage accumulates, signalling through the protein IGF1 is suppressed, and mice age at an accelerated rate. The pathways by which these processes influence ageing are unclear (dashed arrows). **b**, Vermeij *et al.* report that restricting the diets of NER-deficient mice reduces DNA damage and further suppresses IGF signalling. The average lifespan of the mice significantly increases compared with counterparts that eat freely, and they remain healthy for much longer.

dietary-restricted mutants, indicating that the pathways' repression modulates lifespan extension, at least in part.

But how does dietary restriction reduce the accumulation of DNA damage? Although Vermeij *et al.* say it is inconceivable that there is a role for compensatory pathways that enhance DNA repair, it is a speculation that, in our opinion, deserves further research. The authors also speculate that there is an exaggerated response to DNA damage in NER-deficient mice, perhaps as part of an increase in the organism's response to various stress signals. Concomitant adjustments in metabolic regulation, together with alterations in the function of energy-producing organelles called mitochondria, may also shift cellular metabolism towards roles that protect the genome from damage.

Another observation by Vermeij *et al.* that might point towards a mechanism for dietary-restriction-dependent reductions in DNA damage is that molecular stress responses are increased in ERCC1-deficient animals. Such stress responses are modulated, in part, by mTOR signalling⁶. Long-term treatment with rapamycin, a molecule that inhibits mTOR signalling, reduces the accumulation of DNA damage in another genomic-instability disorder, Werner syndrome⁷. There have been other examples of daily rapamycin treatments causing substantial extensions in lifespan — for instance, rapamycin approximately triples the lifespan of mice that lack a mitochondrial protein called *Ndufs4*, which is involved in energy production⁸.

Vermeij and colleagues' study greatly strengthens the evidence supporting the idea that genomic instability is a major mechanism underlying human progeroid syndromes⁹. Moreover, modest dietary restriction could be rapidly and cheaply tested in patients with these conditions. There is little doubt that the authors' findings will lead to peer-reviewed clinical trials of modest dietary restriction, and also probably of mTOR inhibitors, in patients with progeroid syndromes that involve defective DNA repair.

Finally, the study should provide much-needed momentum for efforts to discover pharmacological mimetics of dietary restriction that can be used in humans. But given the enormous genetic and environmental diversity between humans, and the remarkably varied responses of different strains of mice to dietary restriction¹⁰, the responses of individuals to such drugs will probably vary greatly. Large-scale clinical trials will be required before dietary restriction can be recommended as a general treatment for protecting genes during usual ageing. ■

Junko Oshima and George M. Martin are in the Department of Pathology, University of Washington, Seattle, Washington 98195, USA. e-mails: picard@u.washington.edu; gmmartin@u.washington.edu

1. Vermeij, W. P. *et al.* *Nature* **537**, 427–431 (2016).
2. Manandhar, M., Boulware, K. S. & Wood, R. D. *Gene* **569**, 153–161 (2015).
3. Vermeulen, W., Jaeken, J., Jaspers, N. G., Bootsma, D. & Hoeijmakers, J. H. *Am. J. Hum. Genet.* **53**, 185–192 (1993).
4. Niedernhofer, L. J. *et al.* *Nature* **444**, 1038–1043 (2006).
5. Laugel, V. *Mech. Ageing Dev.* **134**, 161–170 (2013).
6. Fontana, L., Partridge, L. & Longo, V. D. *Science* **328**, 321–326 (2010).
7. Saha, B., Cypri, A., Martin, G. M. & Oshima, J. *Ageing Cell* **13**, 573–575 (2014).
8. Johnson, S. C. *et al.* *Science* **342**, 1524–1528 (2013).
9. Hisama, F. M., Oshima, J. & Martin, G. M. *Cold Spring Harb. Perspect. Med.* **6**, a025882 (2016).
10. Liao, C.-Y., Johnson, T. E. & Nelson, J. F. *Exp. Gerontol.* **48**, 1025–1029 (2013).

This article was published online on 24 August 2016.

CORRECTION

The News & Views article “Neuroscience: Flipping the sleep switch” by Stephane Dissel and Paul J. Shaw (*Nature* **536**, 278–280; 2016) incorrectly stated that two potassium-channel proteins modulate a neuronal sleep-to-wake switch in fruit flies. In fact, at least three channels are involved in this process. The News & Views has now been corrected accordingly.

natureINSIGHT



THE PROTEIN
WORLD



Cover illustration
Nik Spencer

Editor, *Nature*
Philip Campbell

Publishing
Richard Hughes

Insights Editor
Ursula Weiss

Production Editor
Elizabeth Batty

Art Editor
Nik Spencer

Sponsorship
Reya Silao

Production
Ian Pope

Marketing
Steven Hurst

Editorial Assistant
Giacomo Russo

The Campus
4 Crinan Street
London N1 9XW, UK
Tel: +44 (0) 20 7833 4000
e: nature@nature.com

**SPRINGER
NATURE**

Our understanding of how proteins fold into distinct shapes and catalyse a broad range of chemical transformations has grown tremendously since the term peptide was coined, more than 100 years ago. Here, we present a collection of reviews that highlight four of the most exciting topics that are being investigated by contemporary protein scientists.

State-of-the-art computational methods can now be used to design non-natural proteins that show promise as therapeutic agents and nanomaterials from scratch. David Baker and his colleagues describe how computers can be used to accurately generate *de novo* designed proteins that assemble into desired, defined shapes.

The proteome — a collection of all proteins that are expressed in a cell in a specific context — is extremely complex, especially given that some proteins exist in several states. Wade Harper and Eric Bennett discuss how cells monitor and regulate the relative abundance of each protein in the proteome through quality control mechanisms. They also explore how this knowledge can be used to improve understanding of the function of cells and to develop potential therapeutic drugs.

Cryo-electron microscopy is in the middle of a revolution (*Nature* 525, 172–174; 2015), and now the technique is being used routinely to obtain high-resolution structures of proteins and protein complexes. Rafael Fernandez-Leiro and Sjors Scheres describe the research that led to this revolution and illustrate how powerful the method has become, especially with respect to membrane proteins and large protein complexes.

The most effective way to perform large-scale measurements of protein complexes and proteomes is to use high-resolution mass spectrometry, with advances in technology enabling almost complete proteomes to be analysed. Ruedi Aebersold and Matthias Mann explain how the technique is being used to catalogue the components of proteomes and their sites of post-translational modification, to identify networks of interacting proteins and to uncover alterations in the proteome that are associated with diseases.

We hope that this collection will inspire the next generation of biochemists, biophysicists and molecular biologists to explore the many uncharted regions of the protein world.

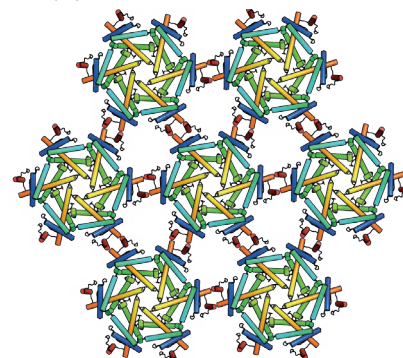
Joshua Finkelstein, Alex Eccleston & Sadaf Shadan
Senior Editors

CONTENTS

REVIEWS

320 The coming of age of *de novo* protein design

Po-Ssu Huang, Scott E. Boyken & David Baker

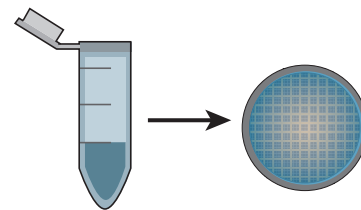


328 Proteome complexity and the forces that drive proteome imbalance

J. Wade Harper & Eric J. Bennett

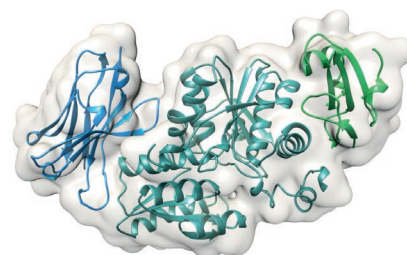
339 Unravelling biological macromolecules with cryo-electron microscopy

Rafael Fernandez-Leiro & Sjors H. W. Scheres



347 Mass-spectrometric exploration of proteome structure and function

Ruedi Aebersold & Matthias Mann



The coming of age of *de novo* protein design

Po-Ssu Huang^{1,2,*†}, Scott E. Boyken^{1,2,3*} & David Baker^{1,2,3}

There are 20^{200} possible amino-acid sequences for a 200-residue protein, of which the natural evolutionary process has sampled only an infinitesimal subset. *De novo* protein design explores the full sequence space, guided by the physical principles that underlie protein folding. Computational methodology has advanced to the point that a wide range of structures can be designed from scratch with atomic-level accuracy. Almost all protein engineering so far has involved the modification of naturally occurring proteins; it should now be possible to design new functional proteins from the ground up to tackle current challenges in biomedicine and nanotechnology.

Proteins mediate the fundamental processes of life, and the beautiful and varied ways in which they do this have been the focus of much biomedical research for the past 50 years. Protein-based materials have the potential to solve a vast array of technical challenges. Functions that naturally occurring proteins mediate include: the use of solar energy to manufacture complex molecules; the ultrasensitive detection of small molecules (olfactory receptors¹) and of light (rhodopsin²); the conversion of pH gradients into chemical bonds (ATP synthase³); and the transformation of chemical energy into work (actin and myosin⁴). Not only are these functions remarkable but they are encoded in sequences of amino acids with extreme economy. Such sequences specify the three-dimensional structure of the proteins, and the spontaneous folding of extended polypeptide chains into these structures is the simplest case of biological self-organization. Despite the advances in technology of the past 100 years, human-made machines cannot compete with the precision of function of proteins at the nanoscale and they cannot be produced by self-assembly. The properties of naturally occurring proteins are even more remarkable when considering that they are essentially accidents of evolution. Instead of a well-thought-out plan to develop a machine to use proton flow to convert ADP to ATP, selective pressure operated on randomly arising variants of primordial proteins, and there were also hundreds of millions of years in which to get it right.

In this Review, we propose that if the fundamentals of protein folding and protein biochemistry and biophysics can be understood, it should become possible to design from the ground up a vast world of customized proteins that could both inform basic knowledge of how proteins work and address many of the important challenges that society faces. We focus specifically on the problem of *de novo* protein design: the generation of new proteins on the basis of physical principles with sequences unrelated to those in nature. We describe the methodological advances that underlie progress in *de novo* protein design as well as provide an overview of the diversity of designed structures for which the high-resolution X-ray crystallography structure or nuclear magnetic resonance (NMR) structure is in atomic agreement with the design model. Almost all protein engineering so far has involved the modification of naturally occurring proteins to tune or alter their function using techniques such as directed evolution^{5–7}, which involves cycles of generating and selecting variation in the laboratory. Because these efforts have been extensively reviewed^{8,9}

and are essentially extensions of evolutionary processes, they will not be discussed here.

It is useful to begin by considering the fraction of protein sequence space that is occupied by naturally occurring proteins (Fig. 1a). The number of distinct sequences that are possible for a protein of typical length is 20^{200} sequences (because each of the protein's 200 residues can be one of 20 amino acids), and the number of distinct proteins that are produced by extant organisms is on the order of 10^{12} . Evidently, evolution has explored only a tiny region of the sequence space that is accessible to proteins. And because evolution proceeds by incremental mutation and selection, naturally occurring proteins are not spread uniformly across the full sequence space; instead, they are clustered tightly into families. The huge space that is unlikely to be sampled during evolution is the arena for *de novo* protein design. Consequently, evolutionary processes are not a good guide for its exploration — as discussed already, they proceed incrementally and at random. Functional folded proteins have been retrieved from random-sequence libraries^{10–12} but this is a laborious (and non-systematic) process. Instead, it should be possible to generate new proteins from scratch on the basis of our understanding of the principles of protein biophysics.

Our approach is built on the hypothesis that proteins fold into the lowest energy states that are accessible to their amino-acid sequences, as originally proposed by Christian Anfinsen¹³. Given a suitably accurate method for computing the energy of a protein chain, as well as methods for sampling the space of possible protein structures and sequences, it should be possible to design sequences that fold into new structures. There are two challenges in implementing this approach: first, the energy of a system cannot be computed with perfect accuracy; and second, the space of possible structures and sequences is very large and therefore difficult to search comprehensively. In this Review, we describe the physical basis for the energy function used in the design calculations and the approaches that are used to overcome the sampling problem. The discussion is based on our experience of developing the Rosetta structure prediction and design methodology¹⁴; other *de novo* protein design software is described elsewhere^{15–17}.

Considerable recent progress in protein design is attributable not only to the advances in understanding and computational methods that are the focus of this Review, but also to advances in two other areas. The first is computing: *de novo* protein design is computationally expensive, and the steady increase in the availability of computing

¹Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA. ²Institute for Protein Design, University of Washington, Seattle, Washington 98195, USA. ³Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA. *These authors contributed equally to this work. †Present address: Department of Bioengineering, Stanford University, Stanford, California 94305, USA.

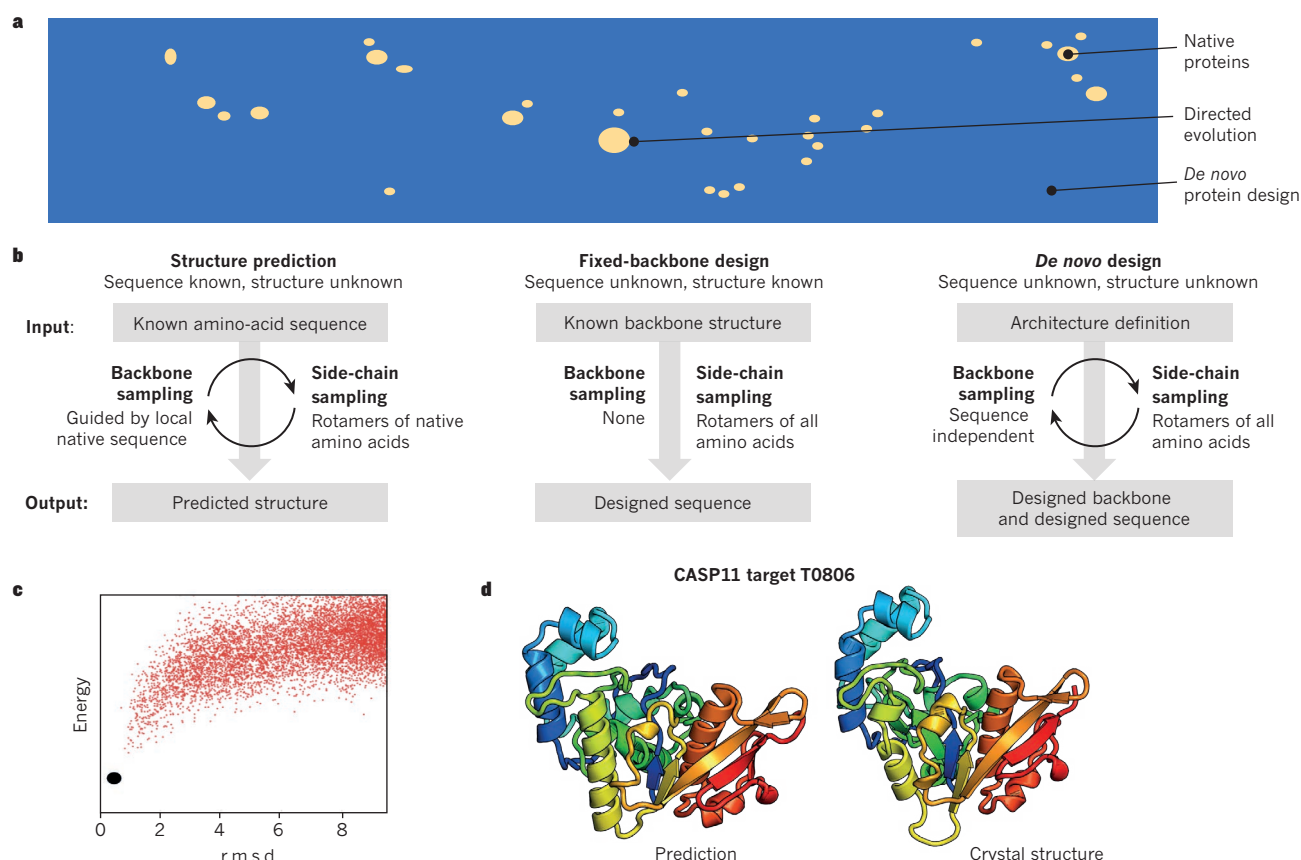


Figure 1 | Methods for *de novo* protein design. **a**, A schematic of the protein sequence space. Evolution has sampled only a tiny fraction of the total possible sequence space (blue), and the incremental nature of evolution results in tightly clustered families of native proteins (beige), which are analogous to archipelagos in a vast sea of unexplored territory. Directed evolution is restricted to the region of sequence space that surrounds native proteins, whereas *de novo* protein design can explore the whole space. **b**, Structure prediction, fixed-backbone design and *de novo* protein design are global optimization problems with the same energy function but different degrees of freedom. In structure prediction, the sequence is fixed and the backbone structure is unknown; in fixed backbone protein design, the sequence is unknown but the structure is fixed; and in *de novo* protein design, neither is known. **c**, Example of an energy landscape generated from fixed-sequence

protein-structure prediction calculations. The red dots represent lowest-energy structures from independent Monte Carlo trajectories, which are plotted according to their similarity to the target structure (black dot) along the *x* axis; structural similarity is measured by root-mean-square deviation (r.m.s.d.). In *de novo* design efforts, designed sequences for which the calculations converge on the target designed structure are selected for experimental characterization. **d**, Blind, *de novo* structure prediction (left) for the critical assessment of protein structure prediction (CASP)11 target T0806, which has no sequence similarity to any protein of known structure, using coevolution-derived contact constraints²⁷. The crystal structure (Protein Data Bank accession code 5CJA) is shown for comparison (right). The ability to predict the structure of proteins with new folds with this level of accuracy enables large-scale structural genomics by means of computer calculation rather than experiment.

power has greatly enabled the work that we describe, much of which was completed using volunteer computing through the Rosetta@home project. The second advance is the synthetic manufacture of DNA. Because the proteins that are being designed do not exist in nature, genes that encode their amino-acid sequences also do not exist. To produce designed proteins in an organism such as *Escherichia coli*, synthetic genes that encode the designed amino-acid sequences must first be manufactured. Methods for DNA synthesis have improved dramatically in the past 10 years, greatly reducing the cost of synthesizing genes for *de novo* designed proteins and increasing the number of computational designs that can be tested experimentally.

Physical principles that underlie protein design

The driving force for protein folding is the burial of hydrophobic residues in the protein's core, away from the solvent. To minimize the size of the cavity that the protein occupies in water, and to maximize van der Waals forces, the side chains in the core must be packed closely but without energetically unfavourable atomic overlaps. Polar groups that interact with the solvent in the unfolded state that become buried upon protein folding must form intra-protein hydrogen bonds to compensate, otherwise the large energy cost of stripping water will disfavour folding¹⁸. The hallmark features of globular protein

structures follow from these considerations: α -helical and β -sheet secondary structures, in which the polar carbonyl and amide groups of the polypeptide backbone can form hydrogen bonds, assemble in such a way that non-polar side chains fit together like the pieces of a jigsaw puzzle to form densely packed cores. Interactions of amino-acid side chains with neighbouring backbone atoms also contribute to the free energy of folding: these include hydrogen bonds at the termini of α -helices and steric and torsional effects that favour certain backbone geometries and disfavour others. For example, the amino acid proline has a rigid internal ring and is compatible with only a narrow range of backbones, whereas glycine, which lacks a side chain, enables tight bending of the backbone in loops between secondary structures.

This picture of protein folding is implemented in an energy function that captures the interactions of the atoms in proteins with each other and with the solvent. The main contributors to this energy function are van der Waals forces that favour close atomic packing, steric repulsion, electrostatic interactions and hydrogen bonds, solvation and the torsion energies of backbone and side-chain bonds. Predicting and designing protein structures using such an energy function requires methods for sampling alternative backbone and side-chain conformations to identify structures and sequences with very low energy. Different methods are used for backbone and side-chain sampling

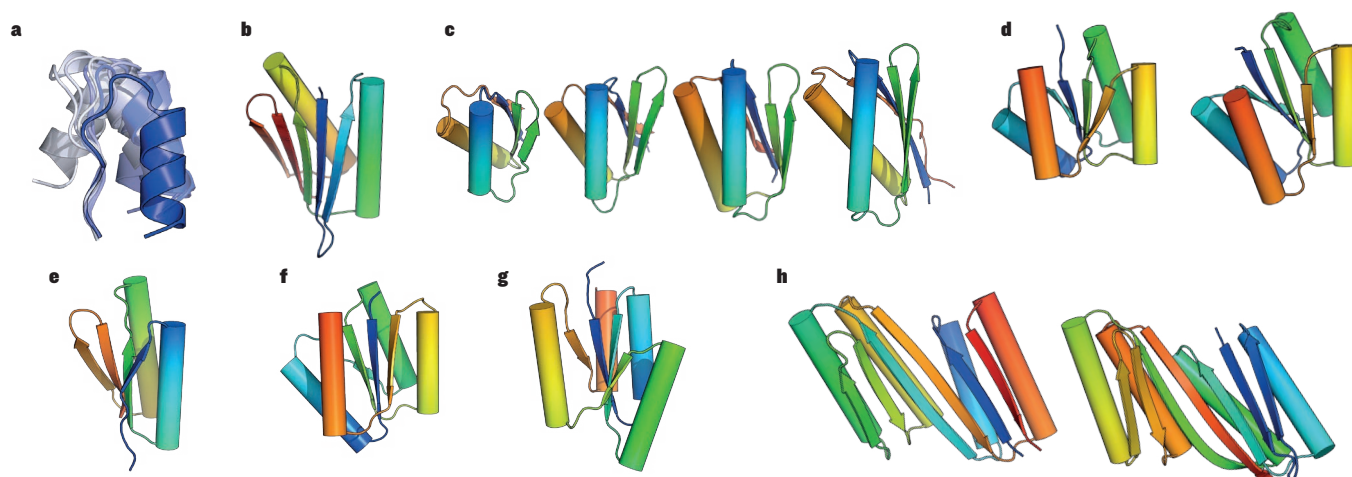


Figure 2 | Designing $\alpha\beta$ proteins. **a**, Sampling alternative backbones for a β -strand-turn- α -helix blueprint through fragment assembly. **b–g**, *De novo* designed ideal $\alpha\beta$ proteins with high-resolution NMR or X-ray structures that are in very close agreement with design models^{28,36,37}. **b**, Top7. **c**, Ferredoxin folds of varying shapes and sizes. **d**, Rossmann 2 \times 2 folds. **e**, IF3-like fold. **f**, P-loop 2 \times 2 fold. **g**, Rossmann 3 \times 1 fold. **h**, Larger, more complex structures that were generated from domains in **b** and **c**³⁸.

(Fig. 1b). In side-chain sampling, discrete combinatorial optimization is used to identify amino acids and side-chain conformations (known as rotamers) that lead to low-energy, closely packed protein cores^{19–22}. If the amino-acid sequence is known in advance, such as in the protein structure prediction problem (predicting the structure of a protein from its amino-acid sequence), the amino-acid identities have already been fixed and the search covers the discrete rotameric states of each side chain. But if the sequence is unknown, such as in the protein design problem (finding a sequence that folds into a specified structure), both the amino-acid identities and the rotameric states are sampled. Backbone sampling often frames the initial stages

of the search as a discrete optimization problem by taking advantage of biases in the local sequence towards a subset of possible local structures. In the later stages of refinement, continuous optimization methods such as quasi-Newton minimization are used to fine-tune the packing and the electrostatic interactions and hydrogen bonding of the structure.

Protein–structure prediction

It is useful to first consider the *ab initio* structure prediction problem: finding the lowest energy structure for fixed amino-acid sequence in the absence of information about the structures of evolutionarily related proteins. Because the amino-acid sequence is fixed, side-chain combinatorial optimization covers only the various rotameric states and the backbone can be built from short fragments with similar local sequences²³. An advantage of this approach is that sampling is very focused in regions where the local sequence strongly favours a particular local structure yet broad in regions where the local sequence is compatible with many conformations. It is still difficult to predict protein structures without homologues of known structure for all but the smallest proteins. The main challenge is the size of the backbone conformational space that must be sampled: the correct structure usually has a lower computed energy than all alternative structures, but it is very hard to find. However, if the sampling is guided by extra sources of information, such as co-evolution-based distance constraints^{24,25}, structure-prediction calculations can find the native-state energy minimum (Fig. 1c). In such cases, accurate, blind predictions of complex protein structures can be made^{26,27} (Fig. 1d).

De novo protein design

Unlike in the structure-prediction and fixed-backbone design problems, in the general (*de novo*) protein design problem, both the sequence and the exact structure of the backbone are unknown (Fig. 1b). Given this, how do we effectively sample backbones from scratch? Because only a small proportion of backbone conformations can accommodate sequences with almost-perfect core packing and hydrogen bonding between the buried hydrogen-bond donors and acceptors, design calculations generally begin with a large set of (more than 10,000) alternative conformations. These initial backbones can be made either by assembling short peptide fragments^{28,29} or by using algebraic equations to specify the geometry parametrically^{30–35}. For each designed backbone conformation, combinatorial sequence-optimization calculations are used to identify the lowest-energy sequence for the structure. *Ab initio* structure-prediction calculations are then carried out to determine whether the designed structure is

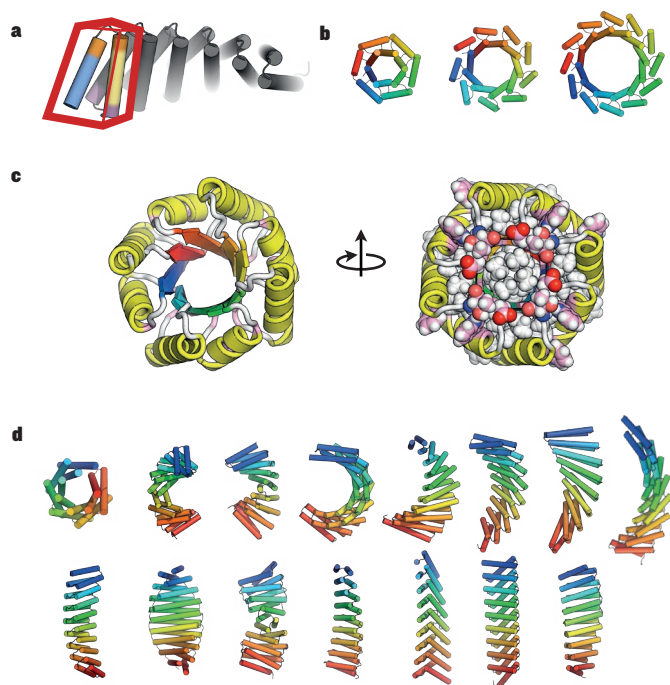


Figure 3 | Designing proteins with internal symmetry. **a**, The propagation of a single repeat unit generates a larger structure. **b–d**, *De novo* designed repeat proteins with high-resolution X-ray structures that are in very close agreement with design models. **b**, *De novo* α -helical toroids⁴¹. **c**, An ideal TIM barrel with four-fold symmetry. Packing features (white) and polar-fold determinants (pink spheres) are shown⁴². **d**, Tandem repeat proteins with a variety of twists and curvatures that go beyond the topologies that are observed in nature⁴³.

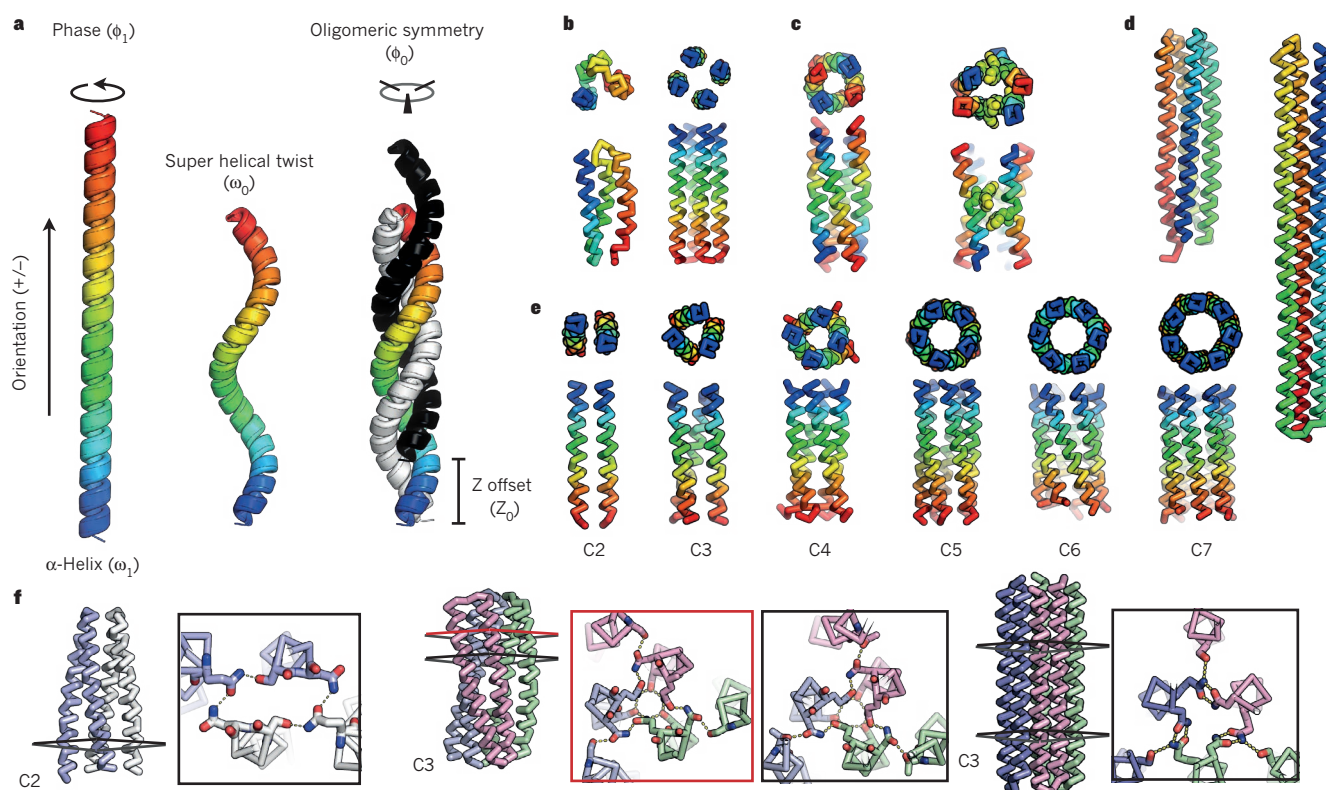


Figure 4 | De novo design using parametric backbone generation.

a, Parameters that describe helical bundle geometry. **b**, The first *de novo* designed helical bundles to be structurally validated: α_3 D (ref. 48) (left) and RH4 (ref. 30) (right), a right-handed coiled coil. **c**, Functional *de novo* helical bundles: a carbon nanotube-binding helix⁵³ (left), and a Zn^{2+} antiporter membrane protein (known as Rocker)³⁴. **d**, Single-chain hyperstable helical bundles³³: a

right-handed four-helix bundle (left) and untwisted three-helix bundles (right). **e**, Homo-oligomeric single-ring helical bundles^{31,33,51,52}. **f**, Homo-oligomeric *de novo* helical hairpins that form double-layered channels with hydrogen-bond network-mediated specificity⁶³; the polar networks are shown as expanded cross-sections. Cn indicates an n -fold cyclic symmetry operation: for example, C2 structures are homodimers and C3 structures are homotrimers.

the lowest-energy state of the designed sequence — this is an important *in silico* consistency check. *De novo* designs are usually experimentally characterized only if structure-prediction calculations that start from the designed sequence strongly converge on the designed structure (Fig. 1c).

Only a finite number of backbones can be sampled computationally. To tackle the important challenge of sequence-independent backbone construction, it is necessary to reduce the enormous space of possible backbone structures to those that are capable of being designed — that is, to those for which there is a reasonable probability that a sequence exists whose lowest-energy state is the structure. Progress towards this goal has required the investigation of sequence-independent constraints on backbone geometry. One such constraint comes from the connectivity of the polypeptide chain and the requirement that the polar atoms of the backbone either make hydrogen bonds within the chain in α -helices or β -sheets or come into contact with the solvent in exposed loops. This constraint immediately restricts the length of the secondary structures that are permitted for a given topology³⁶. Another constraint comes from the limited flexibility of the polypeptide chain, which restricts the lengths of the loops that connect α -helices and β -sheets in various packing orientations³⁷. Simulations and analyses of protein structures have revealed sequence-independent design principles that relate the lengths of helices, strands and loops when packed together that greatly facilitate the construction of topologies that consist of α -helices and β -sheets^{36,37}.

Even with these constraints, the space of possible backbones is still large. To meet the twin goals of bringing the principles that underlie protein folding and structure into sharp focus and generating robust and stable scaffolds for future functional design efforts, much *de novo*

protein-design work has placed an emphasis on designing ideal protein structures with unkinked α -helices and β -strands and minimal loops. By contrast, most naturally occurring proteins contain irregular, non-canonical features that arise either from selection for function or from neutral drift. Such features complicate the structural analysis of proteins and reduce the free energy of folding. (During evolution, there was probably little pressure to optimize the free energy of folding beyond 8 kcal per mol, which corresponds to a folded-state population of more than 99.999%.)

Ideal $\alpha\beta$ folds

A wide range of ideal $\alpha\beta$ protein structures have been designed using the sequence-independent design principles^{36,37} (Fig. 2). The design approach consists of several steps. First, an overall topology ‘blueprint’²⁹ that is consistent with the backbone design principles is created to specify the lengths, packing arrangement and order of the constituent α -helices and β -strands, as well as the lengths of the connecting loops. Second, protein backbones that are compatible with the blueprint are assembled from protein structure fragments using a Monte Carlo approach (Fig. 2a). Third, combinatorial rotamer optimization is used to identify a low-energy amino-acid sequence for each backbone. Fourth, alternating cycles of backbone relaxation and sequence optimization are performed to achieve a sequence–structure pair with very low energy. Last, sequences that converge on the corresponding designed structure in structure prediction calculations are tested experimentally. This design approach was applied to the idealized backbones shown in Fig. 2. Synthetic genes encoding the new designed proteins were generated, and the proteins were produced in *E. coli*. The purified proteins were found to be extremely stable and had structures that were almost identical to those of the design models^{28,36–38} (Fig. 2).

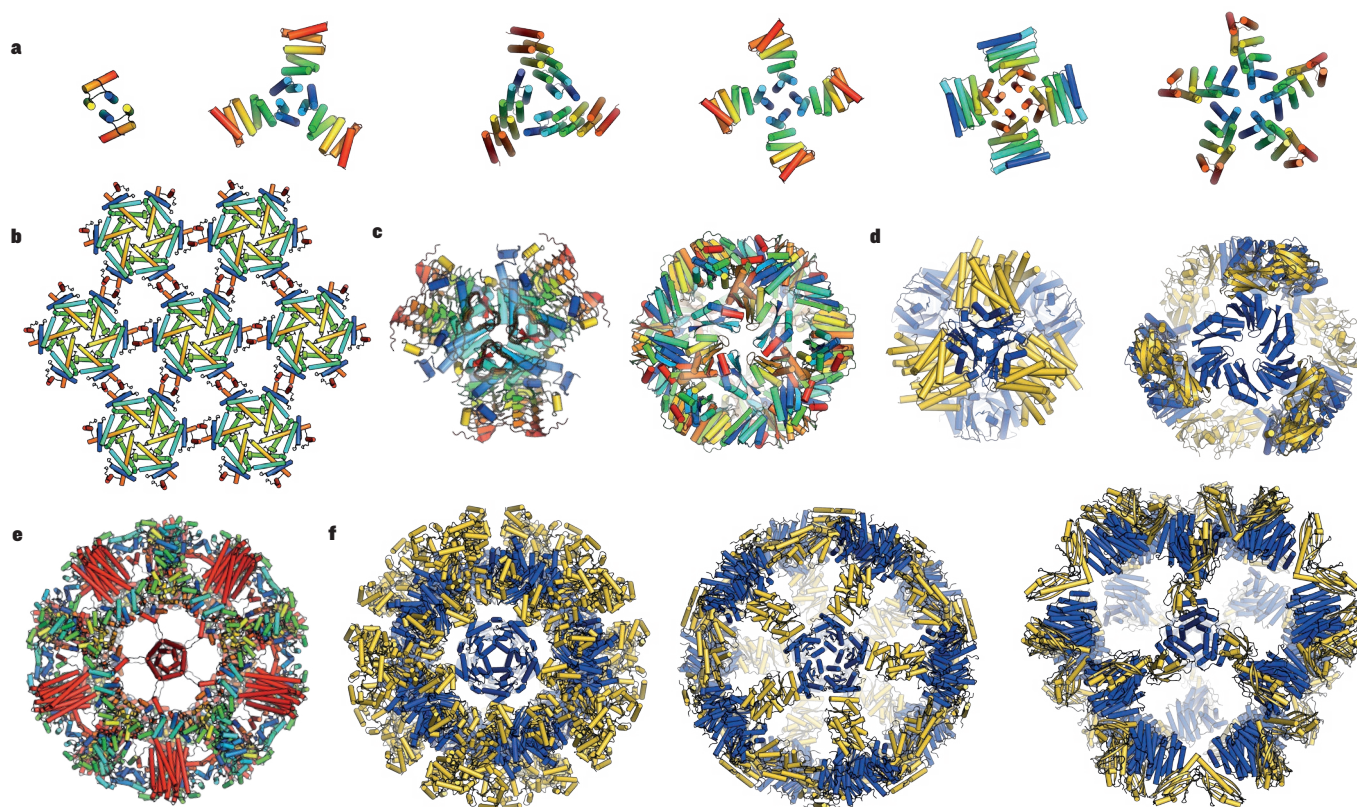


Figure 5 | Designing self-assembling nanomaterials. **a**, C2, C3, C4 and C5 symmetric homo-oligomers (ref. 78 and J. Fallas and G. Ueda, personal communication). **b**, Two-dimensional hexagonal lattice⁸¹. **c–f**, Self-assembling cages. **c**, A one-component tetrahedron (left) and a one-component octahedron⁷⁹ (right). **d**, Two-component tetrahedral

nanoparticles⁸⁰; the two asymmetric components are coloured in blue and yellow. **e**, A one-component hyperstable icosahedron with a *de novo* helical bundle (red helices) fused in the centre⁸². **f**, Two-component megadalton-scale icosahedra⁸³; the two components of each are coloured in blue and yellow.

Repeat proteins

The effort to construct *de novo* proteins with ideal backbone arrangements has led to the design of proteins with internal symmetry in which a single idealized unit is repeated numerous times^{39–41} (Fig. 3). Internal symmetry reduces the size of the sequence space that must be searched and enables a relatively small unit with a known sequence–structure combination to be reused repeatedly to build larger proteins (Fig. 3a). The constraint of internal symmetry is particularly strong for closed structures in which the final repeat unit is juxtaposed with the first, such as in α -helical toroids⁴¹ (Fig. 3b) and the TIM barrel⁴² (Fig. 3c). In the TIM barrel, the backbone design principles, together with the geometry of closed β -sheets, makes four-fold symmetry the highest that can be attained and forces the two α -helices in each α – β – α – β unit to differ in length⁴². Both closed-repeat and open-repeat protein designs have been produced by introducing synthetic genes into *E. coli*, followed by experimental characterization of the purified proteins. High-resolution X-ray crystallography structures for the designs were found to be almost identical to the design models. The α -helical repeat structures have sequences and structures (Fig. 3d) that differ greatly from those found so far in nature, which suggests that naturally occurring proteins sample only a tiny fraction of the stable protein structures that can be realized⁴³. These new repeated proteins are exceptionally stable; several of the open structures are denatured only by guanidine hydrochloride at concentrations of more than 6 M (D. Barrick, personal communication). By contrast, an approach to ‘stitch’ protein structures together from large helix-containing fragments of naturally occurring proteins generates structures with irregularities that are similar to those found in native structures⁴⁴ that present opportunities for the subsequent design of function. Contact information from native structures has also been used to guide the design of new backbone arrangements⁴⁵, including a scaffold that

presents an epitope from respiratory syncytial virus to elicit a neutralizing immune response⁴⁶.

Parametric helical bundles

The use of parametric equations is a complementary approach to generating ideal backbone arrangements that provides considerable control over the global structure. Equations developed by Francis Crick enable the generation of idealized bundles of α -helices in parallel or antiparallel orientations in which the helices have arbitrary lengths, phasing, relative orientations and twists⁴⁷ (Fig. 4a). The helical bundles can be used directly in sequence-design calculations, yielding multiple-subunit oligomeric structures, or the helices can first be connected with loops to yield a single chain. Many helical bundles have been designed in this way^{30,31,33,34,48–52} (Fig. 4), including a peptide that binds to carbon nanotubes⁵³, parallel self-assembling helical channels³¹, an ion transporter³⁴, cages⁵⁴ and an α -helical barrel with installed hydrolytic activity⁵⁵. The combination of parametric backbone generation with combinatorial side-chain optimization has enabled the design of larger, more diverse helical bundles³³; like many *de novo* designed proteins, these parametrically designed proteins are extremely stable, remaining folded in 7 M guanidine hydrochloride at 95 °C.

Hydrogen-bond networks

The principles we have outlined for the *de novo* design of monomeric folds are necessary but not sufficient for controlling the specificity of protein interactions, which despite progress^{56–60} remains a challenge⁶¹. Binding is driven by the balance between the burial of hydrophobic packing residues and peripheral polar interactions that help to solvate the monomeric state and provide structural specificity. In contrast to the double helix of DNA, in which regular arrays of central hydrogen bonds lead to the formation of a high-specificity heterodimer, the hydrogen

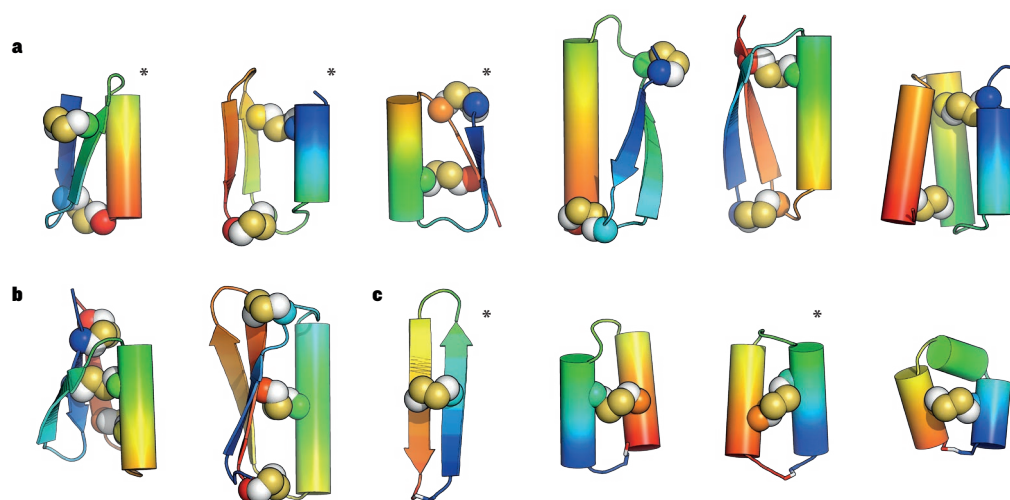


Figure 6 | Designing hyperstable *de novo* constrained peptides. **a, b,** Disulfide crosslinked miniproteins with two (**a**) or three (**b**) disulfide linkages (yellow spheres). **c,** Cyclic peptides with covalently linked N termini and C termini. An asterisk denotes a heterochiral design that contains a mixture of L-amino acids and D-amino acids.

bonds that form at the interfaces of naturally occurring proteins are placed irregularly and are very difficult to design⁶².

A challenge when designing polar interactions is to ensure that all buried hydrogen-bond donors and acceptors form intraprotein hydrogen bonds. In the past year, it has become possible to design with atomic-level accuracy extensive networks of hydrogen bonds in which almost all of the donors and acceptors are satisfied⁶³. This approach has enabled helical-bundle oligomers to be generated with a specificity that is determined by regular arrays of central hydrogen-bond networks, analogous to Watson–Crick base-pairing in DNA⁶⁴. Identification of the rare backbones that can harbour more than one network of hydrogen bonds required the parametric generation of thousands of backbones. In the field of DNA nanotechnology⁶¹, the limited set of Watson–Crick hydrogen bonds has been harnessed to build a wide range of shapes^{65,66}; it should become possible to use similar ‘digital’ design principles to build structures from proteins using modular hydrogen-bond networks to encode specificity.

The design of new functions

The advances described in this Review, most of which were made in the past 3 years, demonstrate that a fundamental understanding of the principles of protein structure and protein folding has been achieved. This knowledge has enabled a wide variety of exceptionally stable protein structures and assemblies to be designed with atomic-level accuracy. (The high-resolution structures for all of the protein designs described in this Review, as determined by NMR, X-ray crystallography or electron microscopy, are in close agreement with the design models.) The potential for designing new functions on the basis of these scaffolds and the more general use of *de novo* backbone design methods is underscored by the achievements of computational protein-design efforts, in which scaffolds from naturally occurring proteins have been repurposed to carry out different functions. Such efforts have yielded enzymes that have attained high catalytic efficiencies through directed evolution^{67–73}, inhibitors of protein–protein interactions that can protect animals from viral infection⁷⁴ and small-molecule binding proteins that can be incorporated into *in vivo* biosensors^{75–77}. The design of precise interfaces between protein subunits has enabled the creation of self-assembling, cyclic homo-oligomers (ref. 78 and J. Fallas and G. Ueda, personal communication), tetrahedra^{79,80}, octahedra⁷⁹ and open two-dimensional assemblies⁸¹ (Fig. 5). Protein interface design methods have been used to create one- or two-component assemblies with icosahedral symmetry and 60 subunits⁸² or 120 subunits⁸³, respectively. The high symmetry of these assemblies enables the multivalent presentation of antigens for vaccine applications, and the large volumes of their interior are well suited to packaging cargo for delivery to targets.

The design of constrained peptides

Because of the level of control that *de novo* protein design offers, the

capabilities of the next generation of designed functional proteins could greatly exceed those of first-generation designed proteins based on native scaffolds. There is also the tremendous potential for *de novo* protein design to go beyond nature to discover new folds by incorporating new chemistries and unnatural amino acids. An example of this is the design of hyperstable peptides, which are constrained by disulfide crosslinks and cyclic peptide linkages that connect the N and C termini⁸⁴. In this case, extensions to the design methodology enabled the use of L-amino acids and D-amino acids within the same protein design (Fig. 6). The structures of these peptides, determined experimentally through NMR and X-ray crystallography, are in close agreement to the design models, and despite the peptides being only 15–50 residues in length, most are extremely resistant to thermal and chemical denaturation.

Improving the robustness of *de novo* design

A limitation of *de novo* protein design is that only a fraction of protein designs adopt stable folded structures when produced in *E. coli*. The most frequent reasons for failure are insolubility and the formation of unintended oligomeric states (polydispersity) — experimentally determined high-resolution structures of soluble and monodisperse designs are almost always very similar to those of the design models. Insolubility and polydispersity probably arise from unanticipated intermolecular hydrophobic interactions. Increasing the robustness of designs will require improvements in the accuracy of the energy function that underlies the design process (for example, explicit modelling of the interactions of protein atoms with specific bound water molecules), more explicit negative design to disfavour alternative states and other advances in computational methodology. As the decreasing cost of synthesizing DNA enables the experimental characterization of larger numbers of protein designs, it should become increasingly possible to identify the features that differ between soluble and insoluble designs. Insight can be obtained by considering the success rate for each class of design that is described in this Review. The highest success rate from the work of our group was obtained for the cyclic and disulfide stapled peptides⁸⁴, for which seven of eight designs were soluble and monodisperse and had structures that were almost identical to the design models; the chemical staples limit alternative conformations of the designs in this class. These designs were also synthesized chemically — the lower success rate for proteins that are expressed recombinantly might be due in part to the toxicity of such proteins in *E. coli* or to other complexities of the bacterium’s biology. The α -helical bundles that are mediated by networks of hydrogen bonds had a solubility of about 90%, and more than 60% of the bundles were monodisperse and in the designed oligomerization state⁶³. Because a large energetic penalty is incurred if buried polar groups do not form hydrogen bonds, altered core-packing arrangements in which hydrogen bonds are not formed are disfavoured. Of the α -helical repeat designs⁴³, 90% were soluble and

64% were monodisperse. Almost all of the monodisperse designs had small-angle X-ray scattering data that were consistent with the design models⁸⁴. Here, the sequence repetition probably favours structures with internal repeats over alternative structures.

Outlook and challenges

A fundamental problem encountered when redesigning naturally occurring proteins to deliver new functions such as catalytic sites is that the alteration of a large number of amino-acid residues to introduce the function will inevitably change aspects of the structure; this is demonstrated by crystal structures of designed enzymes that have unanticipated loop reconfigurations⁸⁵. Native proteins are often marginally stable, and sequence changes can lead to unfolding or aggregation. The very high stability of *de novo* designed proteins should make them more robust starting points for creating new functions.

The next steps in protein design are not without challenges. The ideality of almost all of the *de novo* structures designed so far probably contributes to their stability, and the introduction of functional sites and binding interfaces will inevitably compromise this ideality. Proteins that bind to other proteins usually have hydrophobic residues on their surface and are therefore more prone to aggregation than the idealized polar surfaces of most of the proteins that have been described in this Review, and the active sites of enzymes have some mobility to enable substrates to enter and products to leave. Recessed cavities, which are not incorporated into most *de novo* designed proteins at present, will be required for ligand and substrate binding. Naturally occurring proteins provide numerous examples of the rich functionality, including allostery and signalling, that can emerge in protein systems with multiple low-energy states and moving parts that can be toggled by external stimuli. To achieve such capabilities, which could have widespread applications in the design of molecular machines to tackle problems ranging from tumour recognition to computing, will require proteins to be designed with multiple, distinct energy minima. (By contrast, the *de novo* designs in Figs 2–6 each have a single, deep energy minimum (Fig. 1c).) The creation of a zinc-transporting transmembrane protein that has two alternative states demonstrates that protein design can now start to achieve such complexity³⁴.

Overcoming these challenges in the years ahead is an exciting prospect. Success would signal a technological advance that is analogous to the transition from the Stone Age to the Iron Age. Instead of building new proteins from those that already exist in nature, protein designers can now strive to precisely craft new molecules to solve specific problems — just as modern technology does outside of the realm of biology. ■

Received 12 April; accepted 20 July 2016.

- Firestein, S. How the olfactory system makes sense of scents. *Nature* **413**, 211–218 (2001).
- Rosenbaum, D. M., Rasmussen, S. G. F. & Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **459**, 356–363 (2009).
- Yoshida, M., Muneyuki, E. & Hisabori, T. ATP synthase — a marvellous rotary engine of the cell. *Nature Rev. Mol. Cell Biol.* **2**, 669–677 (2001).
- Spudich, J. A. The myosin swinging cross-bridge model. *Nature Rev. Mol. Cell Biol.* **2**, 387–392 (2001).
- Dougherty, M. J. & Arnold, F. H. Directed evolution: new parts and optimized function. *Curr. Opin. Biotechnol.* **20**, 486–491 (2009).
- Arnold, F. H. The nature of chemical innovation: new enzymes by evolution. *Q. Rev. Biophys.* **48**, 404–410 (2015).
- Goldsmith, M. & Tawfik, D. S. Directed enzyme evolution: beyond the low-hanging fruit. *Curr. Opin. Struct. Biol.* **22**, 406–412 (2012).
- Khouri, G. A., Smadbeck, J., Kieslich, C. A. & Floudas, C. A. Protein folding and *de novo* protein design for biotechnological applications. *Trends Biotechnol.* **32**, 99–109 (2014).
- Regan, L. *et al.* Protein design: past, present, and future. *Biopolymers* **104**, 334–350 (2015).
- Keefe, A. D. & Szostak, J. W. Functional proteins from a random-sequence library. *Nature* **410**, 715–718 (2001).
- Fisher, M. A., McKinley, K. L., Bradley, L. H., Viola, S. R. & Hecht, M. H. *De novo* designed proteins from a library of artificial sequences function in *Escherichia coli* and enable cell growth. *PLoS ONE* **6**, e15364 (2011).
- Murphy, G. S., Greisman, J. B. & Hecht, M. H. *De novo* proteins with life-sustaining functions are structurally dynamic. *J. Mol. Biol.* **428**, 399–411 (2016).
- Epstein, C. J., Goldberger, R. F. & Anfinsen, C. B. The genetic control of tertiary

protein structure: studies with model systems. *Cold Spring Harb. Symp. Quant. Biol.* **28**, 439–449 (1963).

- Leaver-Fay, A. *et al.* ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).
- Wood, C. W. *et al.* CCBUILDER: an interactive web-based tool for building, designing and assessing coiled-coil protein assemblies. *Bioinformatics* **30**, 3029–3035 (2014).
- Negron, C. & Keating, A. E. Multistate protein design using CLEVER and CLASSY. *Methods Enzymol.* **523**, 171–190 (2013).
- Smadbeck, J., Peterson, M. B., Khouri, G. A., Taylor, M. S. & Floudas, C. A. Protein WISDOM: a workbench for *in silico de novo* design of biomolecules. *J. Vis. Exp.* **77**, e50476 (2013).
- Fleming, P. J. & Rose, G. D. Do all backbone polar groups in proteins form hydrogen bonds? *Protein Sci.* **14**, 1911–1917 (2005).
- Ponder, J. W. & Richards, F. M. Tertiary templates for proteins: use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.* **193**, 775–791 (1987).
- Dahiya, B. I. & Mayo, S. L. Protein design automation. *Protein Sci.* **5**, 895–903 (1996).
- Dahiya, B. I. & Mayo, S. L. *De novo* protein design: fully automated sequence selection. *Science* **278**, 82–87 (1997).
- Kuhlman, B. & Baker, D. Native protein sequences are close to optimal for their structures. *Proc. Natl Acad. Sci. USA* **97**, 10383–10388 (2000).
- Simons, K. T., Kooperberg, C., Huang, E. & Baker, D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.* **268**, 209–225 (1997).
- Balakrishnan, S., Kamisetty, H., Carbonell, J. G., Lee, S.-I. & Langmead, C. J. Learning generative models for protein fold families. *Proteins* **79**, 1061–1078 (2011).
- Kamisetty, H., Ovchinnikov, S. & Baker, D. Assessing the utility of coevolution-based residue–residue contact predictions in a sequence- and structure-rich era. *Proc. Natl Acad. Sci. USA* **110**, 15674–15679 (2013).
- Ovchinnikov, S., Kamisetty, H. & Baker, D. Robust and accurate prediction of residue–residue interactions across protein interfaces using evolutionary information. *eLife* **3**, e02030 (2014).
- Ovchinnikov, S. *et al.* Large-scale determination of previously unsolved protein structures using evolutionary information. *eLife* **4**, e09248 (2015).
- Kuhlman, B. Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–1368 (2003).
- Describes Top7, the first globular protein to be designed with a fold not observed in nature.**
- Huang, P.-S. *et al.* RosettaRemodel: a generalized framework for flexible backbone protein design. *PLoS ONE* **6**, e24109 (2011).
- Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T. & Kim, P. S. High-resolution protein design with backbone freedom. *Science* **282**, 1462–1467 (1998).
- Describes RH4, the first protein to be designed using flexible-backbone methods and parametric equations.**
- Thomson, A. R. *et al.* Computational design of water-soluble α -helical barrels. *Science* **346**, 485–488 (2014).
- Grigoryan, G. & DeGrado, W. F. Probing designability via a generalized model of helical bundle geometry. *J. Mol. Biol.* **405**, 1079–1100 (2011).
- Huang, P.-S. *et al.* High thermodynamic stability of parametrically designed helical bundles. *Science* **346**, 481–485 (2014).
- Joh, N. H. *et al.* *De novo* design of a transmembrane Zn^{2+} -transporting four-helix bundle. *Science* **346**, 1520–1524 (2014).
- Presents the design of a functional *de novo* helical bundle (known as Rocker) that can transport Zn^{2+} and Co^{2+} , but not Ca^{2+} , across membranes.**
- Regan, L. & DeGrado, W. F. Characterization of a helical protein designed from first principles. *Science* **241**, 976–978 (1988).
- Lin, Y.-R. *et al.* Control over overall shape and size in *de novo* designed proteins. *Proc. Natl Acad. Sci. USA* **112**, E5478–E5485 (2015).
- Koga, N. *et al.* Principles for designing ideal protein structures. *Nature* **491**, 222–227 (2012).
- Established sequence-independent design principles, which enabled the design of five $\alpha\beta$ topologies.**
- King, I. C. *et al.* Precise assembly of complex beta sheet topologies from *de novo* designed building blocks. *eLife* **4**, e11012 (2015).
- Rämisch, S., Weininger, U., Martinsson, J., Akke, M. & André, I. Computational design of a leucine-rich repeat protein with a predefined geometry. *Proc. Natl Acad. Sci. USA* **111**, 17875–17880 (2014).
- Park, K. *et al.* Control of repeat-protein curvature by computational protein design. *Nature Struct. Mol. Biol.* **22**, 167–174 (2015).
- Doyle, L. *et al.* Rational design of α -helical tandem repeat proteins with closed architectures. *Nature* **528**, 585–588 (2015).
- Huang, P.-S. *et al.* *De novo* design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nature Chem. Biol.* **12**, 29–34 (2016).
- The first structurally verified design of a TIM barrel.**
- Brunette, T. J. *et al.* Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580–584 (2015).
- Jacobs, T. M. *et al.* Design of structurally distinct proteins using strategies inspired by evolution. *Science* **352**, 687–690 (2016).
- Murphy, G. S. *et al.* Increasing sequence diversity with flexible backbone protein design: the complete redesign of a protein hydrophobic core. *Structure* **20**, 1086–1096 (2012).
- Correia, B. E. *et al.* Proof of principle for epitope-focused vaccine design. *Nature* **507**, 201–206 (2014).

47. Crick, F. H. C. The packing of α -helices: simple coiled-coils. *Acta Cryst.* **6**, 689–697 (1953).
48. Walsh, S. T., Cheng, H., Bryson, J. W., Roder, H. & DeGrado, W. F. Solution structure and dynamics of a *de novo* designed three-helix bundle protein. *Proc. Natl Acad. Sci. USA* **96**, 5486–5491 (1999).
49. Fletcher, J. M. *et al.* A basis set of *de novo* coiled-coil peptide oligomers for rational protein design and synthetic biology. *ACS Synth. Biol.* **1**, 240–250 (2012).
50. Zaccai, N. R. *et al.* A *de novo* peptide hexamer with a mutable channel. *Nature Chem. Biol.* **7**, 935–941 (2011).
51. Eisenberg, D. *et al.* The design, synthesis, and crystallization of an α -helical peptide. *Proteins* **1**, 16–22 (1986).
52. Keating, A. E., Malashkevich, V. N., Tidor, B. & Kim, P. S. Side-chain repacking calculations for predicting structures and stabilities of heterodimeric coiled coils. *Proc. Natl Acad. Sci. USA* **98**, 14825–14830 (2001).
53. Grigoryan, G. *et al.* Computational design of virus-like protein assemblies on carbon nanotube surfaces. *Science* **332**, 1071–1076 (2011).
Describes the design of functional helical peptides that coat single-walled carbon nanotubes.
54. Fletcher, J. M. *et al.* Self-assembling cages from coiled-coil peptide modules. *Science* **340**, 595–599 (2013).
55. Burton, A. J., Thomson, A. R., Dawson, W. M., Brady, R. L. & Woolfson, D. N. Installing hydrolytic activity into a completely *de novo* protein framework. *Nature Chem.* <http://dx.doi.org/10.1038/nchem.2555> (2016).
56. Grigoryan, G., Reinke, A. W. & Keating, A. E. Design of protein-interaction specificity gives selective bZIP-binding peptides. *Nature* **458**, 859–864 (2009).
57. Reinke, A. W., Grant, R. A. & Keating, A. E. A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering. *J. Am. Chem. Soc.* **132**, 6025–6031 (2010).
58. London, N. & Ambroggio, X. An accurate binding interaction model in *de novo* computational protein design of interactions: if you build it, they will bind. *J. Struct. Biol.* **185**, 136–146 (2014).
59. Gradišar, H. *et al.* Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. *Nature Chem. Biol.* **9**, 362–366 (2013).
60. Gradišar, H. & Jerala, R. *De novo* design of orthogonal peptide pairs forming parallel coiled-coil heterodimers. *J. Pept. Sci.* **17**, 100–106 (2011).
61. Schreiber, G. & Keating, A. E. Protein binding specificity versus promiscuity. *Curr. Opin. Struct. Biol.* **21**, 50–61 (2011).
62. Stranges, P. B. & Kuhlman, B. A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. *Protein Sci.* **22**, 74–82 (2013).
63. Boyken, S. E. *et al.* *De novo* design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science* **352**, 680–687 (2016).
Describes the design of helical bundles with extensive buried hydrogen-bond networks that mediate interaction specificity in a manner analogous to DNA base pairing.
64. Seeman, N. C. DNA in a material world. *Nature* **421**, 427–431 (2003).
65. Linko, V. & Dietz, H. The enabled state of DNA nanotechnology. *Curr. Opin. Biotechnol.* **24**, 555–561 (2013).
66. Zhang, F., Nangreave, J., Liu, Y. & Yan, H. Structural DNA nanotechnology: state of the art and future perspective. *J. Am. Chem. Soc.* **136**, 11198–11211 (2014).
67. Hilvert, D. Design of protein catalysts. *Annu. Rev. Biochem.* **82**, 447–470 (2013).
68. Kries, H., Blomberg, R. & Hilvert, D. *De novo* enzymes by computational design. *Curr. Opin. Chem. Biol.* **17**, 221–228 (2013).
69. Kiss, G., Çelebi-Ölçüm, N., Moretti, R., Baker, D. & Houk, K. N. Computational enzyme design. *Angew. Chem. Int. Edn Engl.* **52**, 5700–5725 (2013).
70. Richter, F., Leaver-Fay, A., Khare, S. D., Bjelic, S. & Baker, D. *De novo* enzyme design using Rosetta3. *PLoS ONE* **6**, e19230 (2011).
71. Baker, D. An exciting but challenging road ahead for computational enzyme design. *Protein Sci.* **19**, 1817–1819 (2010).
72. Kiss, G., Röthlisberger, D., Baker, D. & Houk, K. N. Evaluation and ranking of enzyme designs. *Protein Sci.* **19**, 1760–1773 (2010).
73. Garrabou, X., Wicky, B. I. & Hilvert, D. Fast Knoevenagel condensations catalyzed by an artificial Schiff-base-forming enzyme. *J. Am. Chem. Soc.* **138**, 6972–6974 (2016).
74. Koday, M. T. *et al.* A computationally designed hemagglutinin stem-binding protein provides *in vivo* protection from influenza independent of a host immune response. *PLoS Pathog.* **12**, e1005409 (2016).
75. Tinberg, C. E. *et al.* Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* **501**, 212–216 (2013).
76. Griss, R. *et al.* Bioluminescent sensor proteins for point-of-care therapeutic drug monitoring. *Nature Chem. Biol.* **10**, 598–603 (2014).
77. Feng, J. *et al.* A general strategy to construct small molecule biosensors in eukaryotes. *eLife* **4**, e10606 (2015).
78. Mou, Y., Huang, P.-S., Hsu, F.-C., Huang, S.-J. & Mayo, S. L. Computational design and experimental verification of a symmetric protein homodimer. *Proc. Natl Acad. Sci. USA* **112**, 10714–10719 (2015).
79. King, N. P. *et al.* Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **336**, 1171–1174 (2012).
80. King, N. P. *et al.* Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510**, 103–108 (2014).
81. Gonen, S., Dimaio, F., Gonen, T. & Baker, D. Design of ordered two-dimensional arrays mediated by noncovalent protein-protein interfaces. *Science* **348**, 1365–1368 (2015).
82. Hsia, Y. *et al.* Design of a hyperstable 60-subunit protein icosahedron. *Nature* **535**, 136–139 (2016).
83. Bale, J. B. *et al.* Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **353**, 389–394 (2016).
84. Bhardwaj, G., Mulligan, V. K., Bahl, C. D. & Baker, D. Accurate *de novo* design of hyperstable constrained peptides. *Nature* <http://dx.doi.org/10.1038/nature19791> (2016).
The design of hyperstable constrained peptides that incorporate both L- and D-amino acids is described.
85. Giger, L. *et al.* Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nature Chem. Biol.* **9**, 494–498 (2013).

Acknowledgements We thank all members of the Baker laboratory and the Institute for Protein Design at the University of Washington, as well as the RosettaCommons community. We apologize to the researchers and protein designers whose work we were unable to acknowledge due to space and scope limitations. The authors are supported by the Howard Hughes Medical Institute (HHMI-027779).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this paper at go.nature.com/2bqnvjlj. Correspondence should be addressed to D.B. (dabaker@uw.edu).

Proteome complexity and the forces that drive proteome imbalance

J. Wade Harper¹ & Eric J. Bennett²

The cellular proteome is a complex microcosm of structural and regulatory networks that requires continuous surveillance and modification to meet the dynamic needs of the cell. It is therefore crucial that the protein flux of the cell remains in balance to ensure proper cell function. Genetic alterations that range from chromosome imbalance to oncogene activation can affect the speed, fidelity and capacity of protein biogenesis and degradation systems, which often results in proteome imbalance. An improved understanding of the causes and consequences of proteome imbalance is helping to reveal how these systems can be targeted to treat diseases such as cancer.

The cellular proteome is exceedingly complex (Box 1). Of the 20,000 or so protein-coding genes of the human genome, a typical cell transcribes about 10,000 genes, resulting in the production of at least as many proteins, which have a cumulative copy number of 10^9 – 10^{11} protein molecules per cell^{1–3}. Although impressive, these numbers fail to demonstrate the true complexity of the cellular proteome due to three important reasons.

First, individual proteins often exist in several modified forms and they also often engage in numerous dynamically regulated protein complexes during their life cycle. For example, large-scale proteomic studies have identified thousands of sites of modification (including sites of phosphorylation, ubiquitylation, methylation and acetylation) in roughly 50% of proteins in humans, the combinatorial nature of which is mostly unknown^{4,5}. It is also estimated that about 100,000 distinct protein isoforms can be generated through alternative splicing from the 20,000 protein-coding genes⁶. The mechanisms that underlie the dynamics, interactions, stoichiometry and turnover of most individual species of protein are poorly understood at the global level (Fig. 1).

Second, cells of different lineages can express distinct sets of genes, including those that promote cellular identity. We are only beginning to understand how differential gene expression in individual cell types translates into differences in the organization and dynamics of the proteome^{7–9}.

Third, variation in the human genome between individuals occurs at the level of around 10^6 differences, a subset of which might alter the abundance and the interactions of the resulting protein products¹⁰. This genomic variation is elevated further in cancer genomes, some of which contain thousands of single base-pair mutations; large-scale chromosomal abnormalities are so prevalent across a range of cancers that about 25% of a typical cancer-cell genome will have undergone a loss or gain in copy number^{11,12}. Our knowledge of how genetic variation alters proteomes in the context of somatic mutations in specific cancers is still in its infancy¹³.

A complete description of the mechanisms that establish cellular proteomes requires not only an understanding of how proteins and their multimeric assemblies are built, but also of the rules that determine how proteins are selected for degradation when they are unable to assemble properly with components of cognate networks. In the past decade, methods have emerged that enable the quantitative analysis of rates of transcription and translation, as well as the determination of relative protein abundance^{2,8,9,14–18}. Moreover, methods for quantifying the

dynamics of protein turnover through the ubiquitin–proteasome and lysosomal systems, which include autophagy, are beginning to yield insights into mechanisms of substrate selection and how these pathways are integrated with stress-response and chaperone networks.

In this Review, we present our knowledge of the cellular systems that monitor and control the abundance and stoichiometry of proteins, as well as the mechanisms of quality control that underlie the cellular response to conditions in which the production and degradation of proteins deviates from the steady state — which we refer to as ‘proteome imbalance’. An emerging understanding of these systems is providing potential avenues for the development of therapeutics that are directed towards diseases of proteome imbalance.

Causes of proteome imbalance

In contrast to the complexity of the genome, we are unlikely to ever know the upper limits of proteome complexity with complete certainty. The protein biogenesis and degradation machineries determine the precise abundance of each protein within the proteome, and both biogenesis and degradation are highly regulated to execute the dynamic control of proteome complexity^{19–21}. The interplay between protein anabolism and catabolism is evident in the consequences of the imbalance in protein homeostasis that can be observed with large-scale genetic variation or an increased demand for ribosomal output (Fig. 1b).

Cancer as a model for proteome imbalance

A number of diverse disorders in people, including many cancers, can be characterized by an imbalance in the proteome that results in chronic proteotoxic stress^{22,23}. Two specific features of cancer cells are considered to contribute to proteome imbalance.

First, the explosion of cancer-genomics data has led to a large catalogue of single base-pair alterations and structural alterations that affect gene copy numbers in numerous types of cancer¹². A potential consequence of this is proteome imbalance caused by mutations that affect the ability of proteins to assemble with cognate complexes or that alter the proteins’ rate of turnover.

Second, sustained proliferation of cells in the absence of growth signals results in deregulated protein production²⁴. In turn, this leads to an elevated demand on the translational apparatus in human cancers. Deviations in the rates of translation are sometimes evident only when translation rates are compared in normal and tumorigenic cells after the loss of permissive growth signals. This loss of growth control is often

¹Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Section of Cell and Developmental Biology, Division of Biological Sciences, University of California, San Diego, La Jolla, California 92093, USA.

achieved through the mutation or genetic amplification of important regulatory proteins, including Myc, phosphoinositide 3-kinases (PI₃Ks) or translation initiation factors such as eIF4E^{25–27}.

The combination of enhanced genetic alteration and increased translational output suggests that tumorigenic cells must acquire an ability to survive despite the persistent generation of an increasingly unstable proteome.

Protein homeostasis and errors in translation

Cellular growth depends absolutely on protein synthesis. The proliferation of unicellular organisms under ideal growth conditions is limited by the speed of translation in such organisms. In *Escherichia coli*, cell division occurs as soon as enough ribosomes have been made to support the growth of another cell^{28,29}. However, rapid protein biogenesis comes at the cost of fidelity. The rate of amino-acid misincorporation in *E. coli* is a single residue every 1,000 to 10,000 amino acids, with lower rates observed in eukaryotic cells^{30,31}. Measurements of error rates for protein biogenesis that take into account mistakes at each step in the messenger RNA translation process are difficult to obtain on a global scale. A conservative estimate provides an overall protein synthesis error rate of 1 in 10,000: for a typical protein of 500 amino acids in length, the synthesis of 1 in every 20 such proteins will deviate from perfect synthesis, and a subset of these defective proteins will be intrinsically unstable. Although this proportion seems to be large at first, the impact of erroneous protein synthesis can be understood only in the context of the cell's capacity for eliminating defective translation products through protein degradation²⁰. Protein half-lives have been determined for a considerable percentage of the proteome using metabolic pulse labelling followed by mass spectrometry. A number of studies report a median protein half-life of 20–46 hours, which reflects an averaged turnover number of the probable most-stable protein isoforms^{2,32,33}. It can be challenging or even impossible to extract quantitative information from these types of studies about the rates of and capacity for proteasomal degradation, or the relative contribution of distinct pools of proteins, including defective translation products, to the overall protein half-life.

An alternative way of estimating the scale of error in protein synthesis and its impact on protein homeostasis is to examine the fraction of total nascent chains that are marked with ubiquitin and targeted for degradation. About 2% of nascent chains in yeast and 12–15% of those in human cells are ubiquitylated^{34,35}. There are between 1 million and 10 million ribosomes in a typical mammalian cell¹⁷; however, the percentage of actively translating ribosomes is unknown in such cells. If only 50% of 5 million cellular ribosomes translate an mRNA that encodes a protein of 500 amino acids at a speed of 5 amino acids per second, 1.5 million proteins will be synthesized every minute³⁶. Under the assumption that 12% of all nascent chains are ubiquitylated and targeted for degradation as the result of an unresolvable error in translation, 180,000 proteins will require ubiquitin-dependent degradation every minute. And assuming a protein degradation rate of 10 amino acids per second (under idealized *in vitro* conditions³⁷) and that there are 1 million proteasomes per cell (ref. 3), 50% of which are active, 600,000 proteins of 500 amino acids in length can be degraded every minute.

These estimates suggest that the defective products of translation are unlikely to burden the ubiquitin–proteasome system, at least under steady-state conditions, and that proteasomes are underloaded. However, careful quantitative studies are needed to more accurately define unknown parameters such as the fraction of active proteasomes in a cell and the *in vivo* rates of turnover for particular pools of individual proteins. The suggestion that there is spare proteasomal capacity in cells is in agreement with cryo-electron microscopy studies that estimate that about 20% of proteasomes in hippocampal neurons are in a substrate-engaged conformation³⁸. And the observation that proteins that are targeted for proteasomal degradation accumulate only after more than 60% of cellular proteasomal activity is inhibited is consistent with the idea that the excess capacity for degradation exists to buffer cells from fluctuations in proteome stress³⁹.

BOX 1

Current knowledge of proteome identity and abundance

The human genome is predicted to encode 20,687 proteins, although its annotation continues to be refined¹³¹. Although RNA sequencing makes it possible to determine which genes in a particular cell type are transcribed, it is more challenging to determine the identity and relative abundance of proteins that are present in that cell type, partly because of the dynamic range of protein abundance and the difficulty in identifying certain types of proteins (such as membrane proteins)¹³². The translation products of 17,294 (ref. 15) and 18,097 (ref. 18) genes have been detected across a wide range of tissues and cell types, including putative translation products from what were previously considered to be non-coding RNAs. Studies in commonly used cell lines have attempted to identify all of the proteins in these cells and some have also tried to compare protein abundance with mRNA abundance^{1–3,8,9,133}. Among the deepest proteomic content measured so far for a single cell line comes from studies of HEK293T cells¹³⁴ and HeLa cells³. In the HeLa cell analysis, in which 10,596 proteins were detected, mRNA transcripts for up to an extra 20% of genes were identified³. Similarly, in the HEK293T cells, in which 10,326 proteins were detected, large-scale interaction proteomic experiments identified 10% more proteins than could be identified through total deep proteomic analysis¹³⁴. Available descriptions of proteomes for such cell lines therefore probably underestimate the total number of proteins that are expressed by 10–20%. Proteomics also enables a means by which to estimate the copy number of individual proteins within a single cell. These estimates range from 1.7×10^{11} proteins to 3×10^9 proteins, depending on the cell type and depth of the analysis^{1–3}. Several features of the HeLa cell proteome demonstrate its complexity¹. For instance, of the 3×10^9 protein molecules per HeLa cell, the top 40 most abundant proteins constitute 25% of the entire proteome by mass, the top 600 most-abundant proteins constitute 75% of the proteome by mass and the least-abundant half of proteins accounts for less than 2% of the proteome by mass. These parameters, albeit incomplete, provide a glimpse into the intricacies of the proteome.

Such model systems do not take into account shifts in the genetic and environmental landscapes that occur over the lifetime of an organism. For example, in mouse models, genetic perturbations of translation-fidelity mechanisms that increase the output of defective translation products result in neurodegenerative phenotypes^{40,41}. The observation that defects in systemic protein quality control often manifest as neurological phenotypes suggests that neurons might be particularly sensitive to conditions that promote proteome imbalance. Indeed, mice with reduced function of the E3 ubiquitin–protein ligase listerin (Ltn1), which is involved in the ubiquitylation of defective nascent chains, have numerous neurological anomalies⁴². These examples hint at the possibility that degradation capacities might be eclipsed, over considerable periods of time, by cellular environments that are permissive to lower-fidelity protein biogenesis.

The response of cells to low-fidelity translation

The balance between the speed and the fidelity of translation has been tuned over time to maximize fitness. Classic studies in *E. coli* have revealed mutations in ribosomal proteins that result in higher-fidelity translation^{31,43,44}. However, this increase in fidelity comes at a

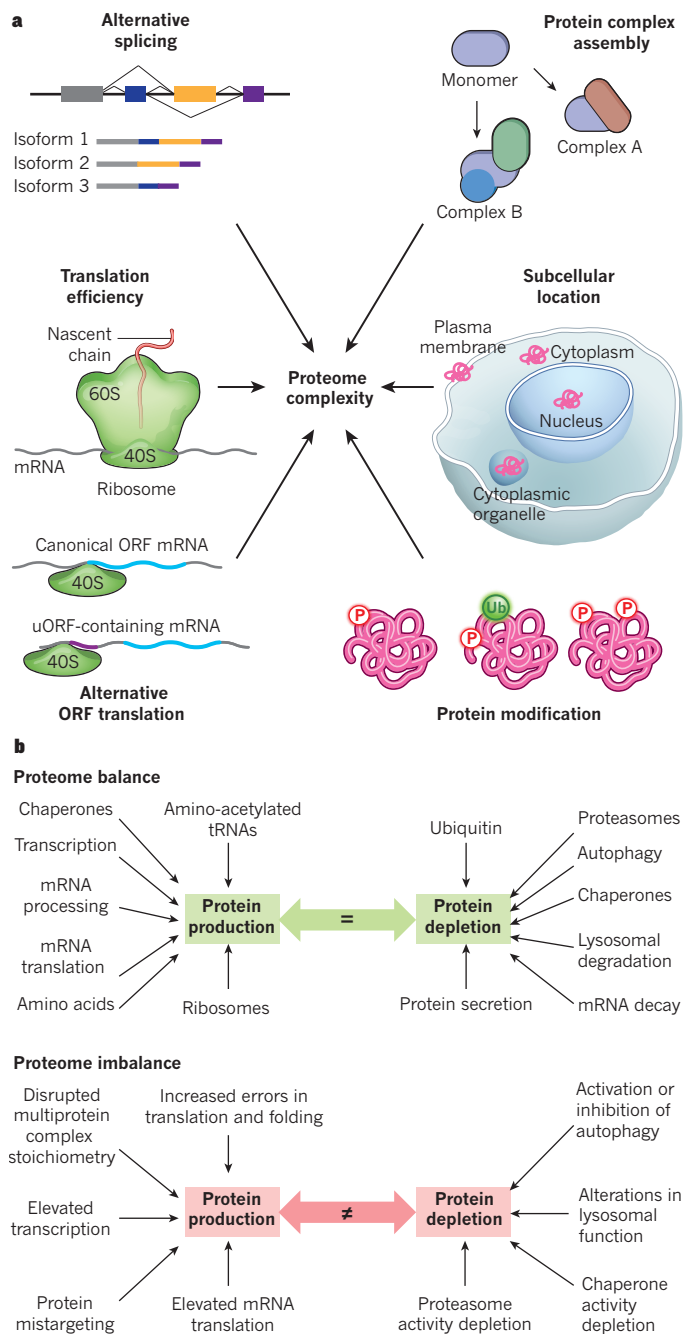


Figure 1 | An overview of proteome complexity. **a**, Numerous factors contribute to the generation of complex proteomes. These include (clockwise from top left): alternative splicing; the assembly of protein complexes with varied compositions; the subcellular location of proteins; the attachment of various modifications to proteins; the use of alternative upstream open reading frames (uORFs, purple) in mRNA translation; and the efficiency of mRNA translation. **b**, In a balanced proteome (green), the level of protein production does not exceed the capacity of protein depletion systems. To achieve this equilibrium, many factors contribute to the generation or degradation of proteins. Cellular events and states such as chromosome imbalance, oncogenic activation and errors in translation alter the proteome in ways that promote proteotoxic stress and lead to imbalance in the proteome (red). This can be buffered by an enhanced capacity for protein degradation or be exacerbated by the depletion of factors that facilitate protein folding or degradation. P, phosphate; Ub, ubiquitin.

considerable fitness cost, as demonstrated by slower rates of growth overall. Decreases in fidelity are also undesirable: in the yeast *Saccharomyces cerevisiae*, the introduction of a transfer RNA that miscodes a

leucine codon as serine results in the chronic mistranslation of the entire transcriptome⁴⁵. This reduction in fidelity produces a transcriptional response that resembles the environmental stress response, as well as a decrease in ribosomal protein mRNAs, a decrease in the overall rates of protein synthesis and an overall reduction in fitness⁴⁵. To survive, cells that have unstable genomes, such as cancer cells, might require compensatory adaptation to such prolonged activation of stress responses. Laboratory evolution experiments performed in yeast that miscode a leucine codon to serine result in rapid adaptation to amino-acid miscoding and the recovery of growth rates⁴⁶. Sequencing has revealed that the adapted strains often contain large-scale genomic deletions and amplifications. The enhanced proteome instability that these genetic changes generate does not seem to be deleterious to the cells, which already have elevated levels of proteome stress. It has been proposed that large-scale chromosomal abnormalities and the sudden increase in proteome stress that they produce force cells to adapt to widespread proteome imbalance⁴⁶. These observations could explain why a large proportion of tumour cells contain chromosomal abnormalities. Aneuploidy may not lead directly to tumorigenesis, but the resulting adaptation to proteome stress could be of benefit to tumour formation.

Ribosome-associated mechanisms of quality control

Errors in translation can occur at numerous points in the process, resulting in the production of defective nascent chains that are removed by ribosome quality control (RQC) pathways. RQC pathways engage stalled ribosomes and catalyse the degradation of the associated nascent chain. This process involves splitting of the 80S ribosome, recruitment of Ltn1, ubiquitylation of the nascent chain, extraction of the nascent chain by the ATPase Cdc48 (the yeast orthologue of p97) and proteasomal turnover of the nascent chain^{47–54}. Considerable insight into this quality control system and a potential signalling arm of the pathway came from the discovery of Rqc2, a component of the complex that is responsible for RQC, and its role in adding C-terminal alanine and threonine extensions (known as CAT tails) to stalled nascent chains that cannot be ubiquitylated (for example, in Ltn1-mutant cells)⁵⁵. Because Rqc2 was discovered in a screen for activators of heat shock factor protein 1 (HSF1)⁵⁶, it is possible that the addition of CAT tails might directly or indirectly signal heat-shock activation, although further studies are needed to define the underlying mechanisms. A role for Rqc2, Ltn1 and CAT-tail formation in mediating the aggregation and toxicity of both stall-inducing nascent chains and mutated Huntingtin that contains expanded and pathogenic polyglutamine repeats has been demonstrated in *S. cerevisiae*^{57–59}. Despite the elegant biochemical characterization of the RQC complex in mammalian cells, a physiological role for the complex or CAT-tail formation in mammals has yet to be demonstrated, and endogenous Ltn1 substrates are lacking. The abundance of ribosomes is about 300 times greater than that of Ltn1 (ref. 17), which suggests that the capacity of Ltn1 and the function of the RQC complex could be easily overcome by small perturbations in the proteome, or that alternative factors and pathways also function in regulating RQC. Ribosome stalling that is induced by translation elongation inhibitors has been shown to stimulate the site-specific regulatory ubiquitylation of particular 40S ribosomal proteins; it is therefore possible that post-translational mechanisms contribute to sensing or clearing stalled ribosomal complexes⁶⁰. However, a direct role for these ubiquitylation events in mediating RQC has yet to be determined.

Another important consideration is the physiological conditions under which ribosome stalling occurs at high enough levels to elicit an adaptive response. Such conditions might include the limited availability of particular amino acids. Ribosome profiling studies in cells that were treated with L-asparaginase, which converts asparagine to aspartic acid, revealed the robust accumulation of ribosomes that were stalled on asparagine codons⁶¹. And in patient-derived clear cell renal cell carcinoma tumours, ribosomes were found to stall at proline codons. The same tumours showed an increase in a crucial proline biosynthetic enzyme as part of a possible compensatory feedback loop⁶¹. Together,

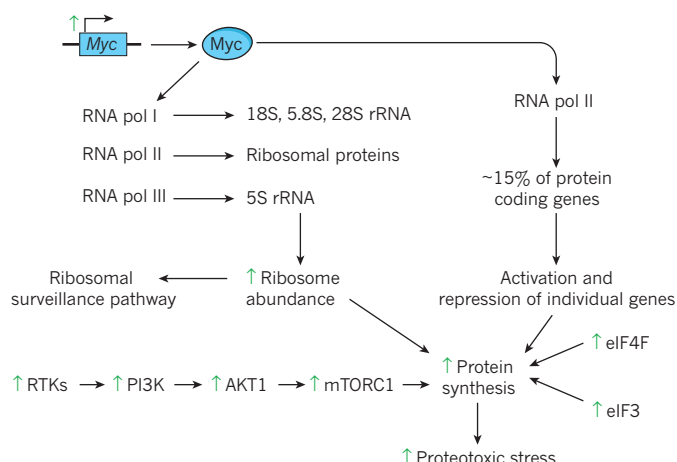
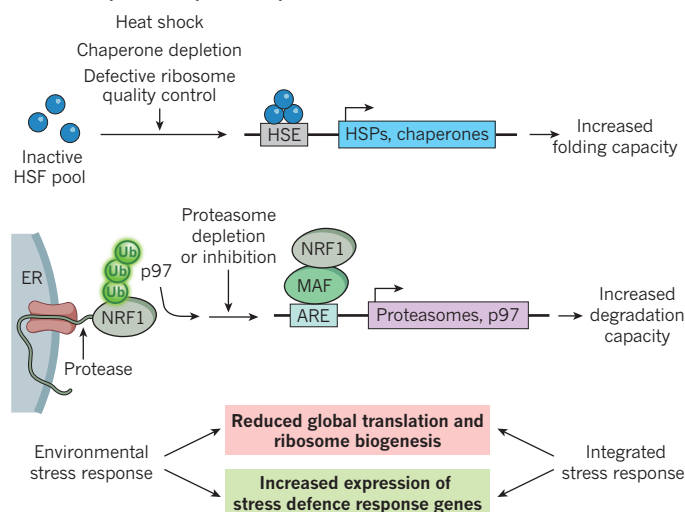
a Genetic alterations that drive proteome imbalance**b Transcriptional responses to proteome imbalance**

Figure 2 | Mechanisms that contribute to proteome imbalance and transcriptional responses. **a**, Genetic alterations in human cancers can generate proteome imbalance. Myc protein promotes the RNA polymerase (RNA pol) I- and RNA pol III-dependent transcription of rRNAs; it also promotes the RNA pol II-dependent transcription of ribosomal proteins to induce ribosome production. Elevated levels of some ribosomal proteins in excess of the assembled ribosome can lead to the activation of the ribosomal surveillance pathway, which stimulates cell death pathways. At the same time, Myc activates about 15% of protein-coding genes that are transcribed by RNA pol II, thereby increasing the protein-synthesis load of the cell. This is likely to increase the number of defective translation products, which lead to an increase in proteotoxic stress. Similarly, the activation of receptor tyrosine kinases (RTKs) and the PI₃K signalling pathways that stimulate AKT1 and mTORC1 activity, or the overexpression of translation initiation factors eIF4F and eIF3 can increase protein synthesis, also leading to increased levels of proteotoxic stress. **b**, Two transcriptional response systems sense imbalance in the proteome and can elevate the cell's capacity for protein

folding or degradation. One mode of control involves the conversion of a pool of inactive HSF proteins in response to perturbations such as heat shock into an active nuclear transcription complex that binds a promoter known as a heat-shock element (HSE). This leads to the expression of heat-shock proteins (HSPs), including chaperones, and increases the cell's capacity for protein folding. Another mode of control is the production of proteasome subunits to increase the cellular capacity for protein degradation. Under basal conditions, transcription factor NRF1 in the endoplasmic reticulum (ER) is targeted for p97-dependent proteasomal degradation through ER-associated degradation (ERAD) (not shown). However, when the activity of the proteasome is inhibited or depleted, retrotranslocated NRF1 is cleaved by an unknown protease, which facilitates the translocation of NRF1 into the nucleus where it activates the transcription of proteasome subunit genes. Global environmental and integrated stress-response pathways also attempt to rebalance the proteome through reduced translational output (red box) and enhanced cellular defence systems (green box) such as protein folding and degradation machineries. ARE, antioxidant response element; MAF, transcription factor MAF.

these results suggest that altered amino-acid metabolism in some cancers might result in enhanced ribosome stalling. Whether these cancers display an increased need for RQC has yet to be examined.

Alterations in translational output

Reoccurring tumorigenic mutations in diverse genes stimulate or sustain mRNA translation under conditions that would otherwise dampen protein output^{24,26}. One such mutation is the amplification of the oncogene *Myc*, which occurs in 50% of human cancers²⁵. Myc functions as a transcription factor and it is estimated that the transcription of 15% of human genes is stimulated by the presence of activated Myc⁶². Myc has the ability to stimulate not only the production of ribosomal proteins but also ribosomal RNA synthesis by activating all three RNA polymerases^{25,62} (Fig. 2a), which directly leads to an increase in ribosome biogenesis and translational capacity. The global increases in ribosome biogenesis and the overall capacity for translation that are observed on Myc activation are thought to facilitate tumorigenesis. In support of this, the loss of translational capacity that results from the deletion of a single copy of the ribosomal protein gene *RPL24* limits tumorigenesis in a widely used mouse model of B-cell lymphoma that is driven by Myc overproduction⁶³. This result provides a clear example in which rebalancing the proteome stifles the development of cancer.

The activation of cellular signalling pathways that stimulate either translation initiation or elongation is common among tumorigenic cells. This activation is achieved often through mutations that stimulate PI₃K signalling²⁷. Such mutations can occur in upstream receptor tyrosine kinases, in PI₃Ks themselves or in downstream effectors such as the kinase Akt1, with each resulting in the general, sustained activation of the kinase-containing signalling complex mTORC1 (refs 24 and 27).

Although mTORC1 signalling impinges on a large array of cellular systems, it directly stimulates mRNA translation through activating phosphorylation of the ribosomal protein S6 kinase β -1 (RPS6KB1) and inhibitory phosphorylation of eIF4E binding protein 1 (EIF4EBP1) (ref. 24). The sustained hyperactivation of translation initiation and elongation that is observed on mTORC1 activation might lead to a further loss of fidelity in mRNA translation and to the increased production of defective products. Indeed, cells with chronically activated mTORC1 show a decrease in translational fidelity, which results in an enhanced sensitivity to proteotoxic stress⁶⁴. This suggests that tumorigenic cells with mutations that activate translation must adapt to a heightened basal level of proteome imbalance. Accordingly, cells that are exposed to proteasome inhibitors respond by increasing the transcription of genes that encode proteasome components through activation of the transcription factor NRF1 (refs 65–67) (Fig. 2b). The disruption of NRF1 in cancer might therefore limit their ability to adapt to proteome imbalance.

Another mechanism by which cancer cells enhance translation is the overproduction of translation initiation factors^{22,26}. Components of both the eIF3 and eIF4F initiation complexes are overexpressed in a wide range of cancers^{22,26}. The idea that enhanced translation initiation can promote tumour formation is best described for eIF4E, the mRNA 5'-cap-binding component of the eIF4F complex. Overexpression of eIF4E mediates both cellular transformation in cell-culture models and enhanced susceptibility to tumours in mouse models^{68–70}. The ability of eIF4E to stimulate the initiation of translation is dependent on mTORC1 activity as the loss of inhibitory EIF4EBP1 phosphorylation sequesters eIF4E from the eIF4F complex⁷¹. Interestingly, the overexpression of a mutant EIF4EBP1 that evades inhibitory mTORC1 phosphorylation

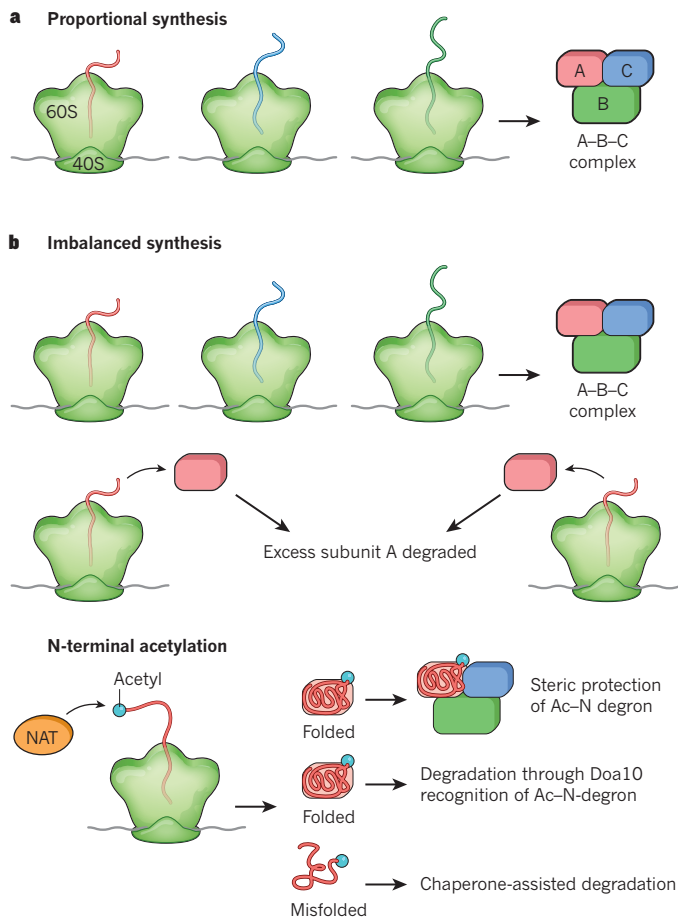


Figure 3 | Regulating the stoichiometry of protein complexes. **a**, In proportional synthesis, the abundance and translation rates of mRNA are tuned to produce the appropriate stoichiometry for the formation of multiprotein complexes with the subunits A, B and C. **b**, In imbalanced synthesis, subunit A is expressed at a higher rate than subunits B and C (top), either as a result of increased transcription and translation via gene dosage effects or oncogene activation, or through genetic programming. Excess subunit A is then degraded by the ubiquitin–proteasome system. N-terminal acetylation (blue circle) through the N-acetyltransferase (NAT) system (bottom) is also proposed as a mechanism by which supernumerary subunits are marked for degradation. After the N-acetylated protein is assembled into a complex, the N-terminal acetyl residue (Ac–N degnon) can no longer be recognised by the E3 ligase Doa10 due to steric blockage, and the protein is not degraded. Folded N-acetylated proteins that have not been incorporated into complexes are detected by Doa10 and marked for degradation, and misfolded proteins that contain Ac–N degnons are degraded with assistance from chaperone proteins.

or a 50% reduction in the levels of eIF4E does not lead to global alterations in basal translation^{72,73}. However, the expression of mutant EIF4EBP1 blocks eIF4-driven tumorigenesis in haematological cancers and prostate cancer in mice, and mice with reduced levels of eIF4E show suppressed development of GTPase KRAS-driven lung cancers^{71,73,74}. Together, these results indicate that the suppression of hyperactivated translation can inhibit tumorigenesis. However, it remains unknown how the enhanced translation in these cases affects protein degradation systems, and specifically RQC, or whether adapting to the increased demand for protein degradation is a necessary step for tumorigenesis.

Changes in the stoichiometry of multisubunit complexes

Keeping the proteome in balance requires more than just quality control processes at the ribosome: it also involves mechanisms that control the stoichiometry of complexes. Imbalance in subunit stoichiometry can have detrimental effects by disturbing the assembly or dynamics

of multisubunit complexes. This is demonstrated by the α – β -tubulin complex, which forms dynamic filaments that are crucial for cell division and organelle trafficking. In classic experiments, the 1.4-fold overexpression of β -tubulin leads to the disassembly of microtubules, the formation of alternative β -tubulin-positive structures and a loss of cell viability, which can be rescued by the expression of α -tubulin^{75,76}. This phenotype reflects the sequestration of factors that are necessary for the formation of α – β -heterodimers by elevated β -tubulin. By contrast, the 30-fold overexpression of α -tubulin has no obvious phenotype^{75,76}. This example reinforces the idea that cells require mechanisms for controlling the stoichiometry of multisubunit complexes. Proportional synthesis through tuned efficiencies of transcription and translation probably controls the stoichiometry of some complexes⁷⁷ (Fig. 3a). However, turnover of excess or orphan proteins through the proteasome or the lysosome also plays a considerable part in coupling protein abundance with the assembly of complexes.

The response of cells to increased gene dosage

Our understanding of how cells respond to changes in protein abundance has been enhanced by examining the consequences of large-scale changes in chromosome copy number on the proteomes of *S. cerevisiae*, as well as in plants and mammalian cells, to a lesser extent. The analysis of haploid *S. cerevisiae* cells into which 13 of the 16 yeast chromosomes were individually introduced to produce a ‘disome’ has revealed several common principles by which cells respond to proteome imbalance^{78–82} (Fig. 1a, b). In a manner that is largely independent of the particular chromosomes that are duplicated, a two fold increase in transcription from the supernumerary chromosome triggers a chain of events that decrease the fitness of the cell and reduce the cell’s ability to cope with proteome imbalance⁸³. Most evidence suggests that reduced fitness and the presence of proteotoxic stress do not reflect the expression of particular proteins from the disome⁸⁴. Instead, they are a reflection of a greater reliance on protein chaperones and degradation machinery within the cell that attempts to manage the higher levels of protein production, the ensuing increase in translational errors that give rise to misfolded proteins, the imbalance in subunits of multiprotein complexes that accompanies enhanced expression from the disome and a reduction in the capacity of chaperones^{83,85} (Fig. 1b).

Although transcription from the disome produced a two fold increase in the abundance of mRNA, protein abundance diverged from a normal distribution, and the abundance of about 20% of disome-derived proteins increased by only 1.6-fold, compared to a two fold increase in mRNA^{78,82}. Interestingly, this group of disome-derived proteins was enriched in those that participate in multiprotein complexes⁸², which implies that protein degradation systems might participate in the maintenance of the stoichiometry of complexes. Although some proteins might be ‘immune’ to post-translational copy number control, others are therefore regulated by a form of dosage compensation that probably balances the abundance of subunits for multiprotein complexes. Interestingly, a transcriptional response that is related to the environmental stress response^{82,86} seen in cells that are exposed to external stress agents⁸⁷ is layered on top of post-translational control (Fig. 2b). The environmental stress response is observed widely in cells from aneuploid fission yeast, mice, humans and plants, which implies that it is a conserved transcriptional response to proteome imbalance⁸⁶.

Two lines of evidence suggest that a reduction in the dosage of disome-derived proteins can promote cellular fitness. First, deletion of the proteasome-associated deubiquitylating enzyme Ubp6 — a negative regulator of protein turnover — reverses the slow-growth phenotype of disomic cells^{82,88,89}. Proteomic analysis hints that this effect is a reflection of the enhanced turnover of relatively abundant proteins, which reduces proteotoxic stress to promote fitness, although the removal of specific toxic proteins cannot be ruled out. Second, diploid cells that receive an extra chromosome are less sensitive to the perturbation than haploid cells, which suggests that excess protein derived from the supernumerary chromosome is more easily buffered under these conditions^{79,83}.

Genetic screens in *S. cerevisiae* for genes that increase the viability of disomic cells have led to the identification of mutant alleles in several genes that are related to protein homeostasis⁸⁶. These include genes that encode the proteasome subunit Rpt1, two ubiquitin ligases (Rsp5 and Ubr1), Lad2 (the yeast orthologue of the cullin–RING E3 regulator CAND1), a protein involved in rRNA production called Utp1 and a protein involved in vacuolar targeting that is known as Vps64, indicating that there are other mechanisms by which cellular defects in protein imbalance can be ameliorated. Further mechanistic studies are needed to define the underlying pathways. However, the spectrum of genes that have been identified so far suggest that mechanisms that reduce flux through the ubiquitin–proteasome system probably contribute to the toxic effects of aneuploidy in disomic yeast cells.

Demands on the cellular machinery

Widespread proteome imbalance that arises from supernumerary chromosomes has two main effects: the loss of the buffering capacity of protein-folding chaperones and the degradation of proteins that are unable to join cognate complexes (Fig. 3b, top). Although a sizable proportion of the proteome is degraded as a result of either proteome imbalance or errors in translation or folding, we do not yet have a systematic understanding of the rules that dictate the turnover of excess subunits of complexes. This is partly because of a lack of methods that measure the ubiquitylation and turnover dynamics of proteins that are unable to assemble properly into complexes, or that measure rates of turnover for the same proteins in distinct complexes or subcellular compartments.

What are the mechanisms by which proteins that are unable to assemble with their partners are marked for degradation? N-terminal acetylation and a variation of the N-end rule⁹⁰ have been implicated in one such mechanism (Fig. 3b, bottom). This pathway involves the E3 ligase Doa10, which recognizes acetylated N-terminal residues (or Ac–N-degrons) in target proteins⁹¹. Because N-terminal acetylation typically occurs at the same time as translation, the model predicts that newly synthesized proteins will be marked with this signal during their generation. If such proteins are incorporated successfully into cognate complexes, the Ac–N-degrons will be masked by other subunits in the complexes. But if these proteins remain unassembled, they are ubiquitylated by Doa10 and are then degraded by the proteasome. This mechanism has been demonstrated in *S. cerevisiae* for the proteins Cog1 (part of the conserved oligomeric Golgi complex) and Hcn1 (part of the anaphase-promoting complex)⁹². N-terminal proteomics studies suggest that about 85% of detected human proteins are at least partially N-acetylated^{93,94}. Although such studies do not analyse the N terminus of every protein in a sample, the high degree of N-acetylation observed indicates that the Ac–N-degron recognition signal might not be universal because most soluble proteins that do not form tight stoichiometric complexes would be expected to be destroyed. However, a more careful analysis of the N termini of both monomeric proteins and proteins that participate in complexes is needed. N-acetyl groups can also mediate functional interactions with complexes in a dynamic setting without targeting the N-acetylated protein for degradation⁹⁵. A bipartite signal would therefore be the minimum requirement for recognizing orphan proteins that have been stranded from their complexes. One possibility involves the use of hydrophobic surfaces as a second signal that could be recognized by protein chaperones. Indeed, orphan subunits of fatty acid synthase (FASN) as well as orphan von Hippel–Lindau disease tumour suppressor are recognized by the Hsp40–70–90 heat-shock protein system, and FASN is known to be directed to Ubr1, which is a principal quality control ligase^{96–98}. FASN is also one of the most highly ubiquitin-modified proteins by abundance, as are other proteins that are obligate members of complexes, which indicates that this pathway is active under steady-state conditions and that these substrates would be the first to accumulate on proteasome inhibition⁹⁹. Excess components of multisubunit complexes could offer an abundant source of degradation substrates that greatly exceed the amount of substrates that arise from errors in translation.

An important example that has emerged from the 13 yeast disome strains is the dosage compensation of essentially all detected ribosomal proteins⁸². Considering the cellular resources that are used in ribosome production, it is unsurprising that cells actively control the stoichiometry of ribosomes. Early experiments in mammalian cells demonstrated that a disruption in the synthesis of rRNA or in the overproduction of an individual ribosome subunit had little impact on total ribosomal-protein synthesis but that it did result in the turnover of ribosomal proteins through unknown mechanisms^{100,101}. In *S. cerevisiae*, several ribosomal proteins do not accumulate when overexpressed, and these excess subunits are degraded by the proteasome, seemingly without the contribution of autophagy¹⁰². Unlike ribosomal proteins that have been assembled into subunits, which are mainly found in the cytoplasm under steady-state conditions, excess ribosomal protein Rpl26a accumulates in the nucleus. Interestingly, when the proteasome is inhibited, endogenous newly synthesized ribosomal proteins aggregate in the nucleus, suggesting that an important quality control mechanism is present to ensure that the proper stoichiometry of ribosomal proteins is maintained¹⁰². Understanding the machinery that is involved in this quality control process will be a crucial next step. Intriguingly, a subset of extra-ribosomal proteins participates in the ribosome surveillance pathway in which excess ribosomal proteins RPL11 and RPL5 physically inhibit the p53 ubiquitin ligase Mdm2 (refs 103 and 104), thereby promoting p53-dependent cell death. However, it is still unknown how quality control pathways can distinguish this subset of ribosomal proteins from others that are to be degraded rapidly.

Mislocalized proteins

The stoichiometry of protein complexes is only one of several types of alterations that can promote imbalance in the proteome. For example, newly synthesized membrane proteins occasionally fail to translocate properly into the endoplasmic reticulum, which results in their aggregation in the cytoplasm through hydrophobic transmembrane domains. Proteins that contain transmembrane domains can be recognized by a surveillance complex containing the proteins BAG6, TRC35 (also known as GET4), UBL4A and SGTA^{105–107}, which then facilitates the ubiquitin-dependent degradation of the transmembrane client protein through the ubiquitin ligase RNF126 (ref. 108). A distinct system that involves members of the ubiquilin protein family has been identified for proteins containing a C-terminal transmembrane sequence that fails to insert into the mitochondrial outer membrane¹⁰⁹. Ubiquilins (UBQLN1, UBQLN2, UBQLN3 and UBQLN4 in humans) contain an N-terminal ubiquitin-like domain, a central domain that is related to a chaperone-binding domain from heat-shock protein STI1 that associates with hydrophobic sequences, and a C-terminal ubiquitin-associated domain, which binds ubiquitin. The association of client proteins with ubiquilins promotes the ubiquitylation of client proteins by an unknown E3 ligase as well as rapid proteasomal degradation that is mediated by the association of ubiquilin's ubiquitin-like domain with the proteasome¹⁰⁹. The impairment of mitochondrial import or the overproduction of mislocalized mitochondrial proteins initiates a cytoplasmic stress response that includes activation of the proteasome and a reduction in translation¹¹⁰. Similarly to when cells are exposed to stress through chronic mistranslation, cells that experience high levels of mislocalized proteins adapt through enhancements to protein-degradation systems⁴⁶. Further studies are needed to understand the global role of such membrane-protein quality control pathways and to elucidate the contributions of membrane proteins to proteome imbalance.

Autophagy and the control of proteome imbalance

Autophagy is a catabolic process in which proteins and cellular organelles are targeted to the autophagosome for delivery to the lysosome, a site of degradation. This process involves the lipidation of ubiquitin-like ATG8 proteins (MAP1LC3A, MAP1LC3B and MAP1LC3C and GABARAP, GABARAPL1 and GABARAPL2 in mammalian cells) to promote autophagosome formation, maturation and cargo

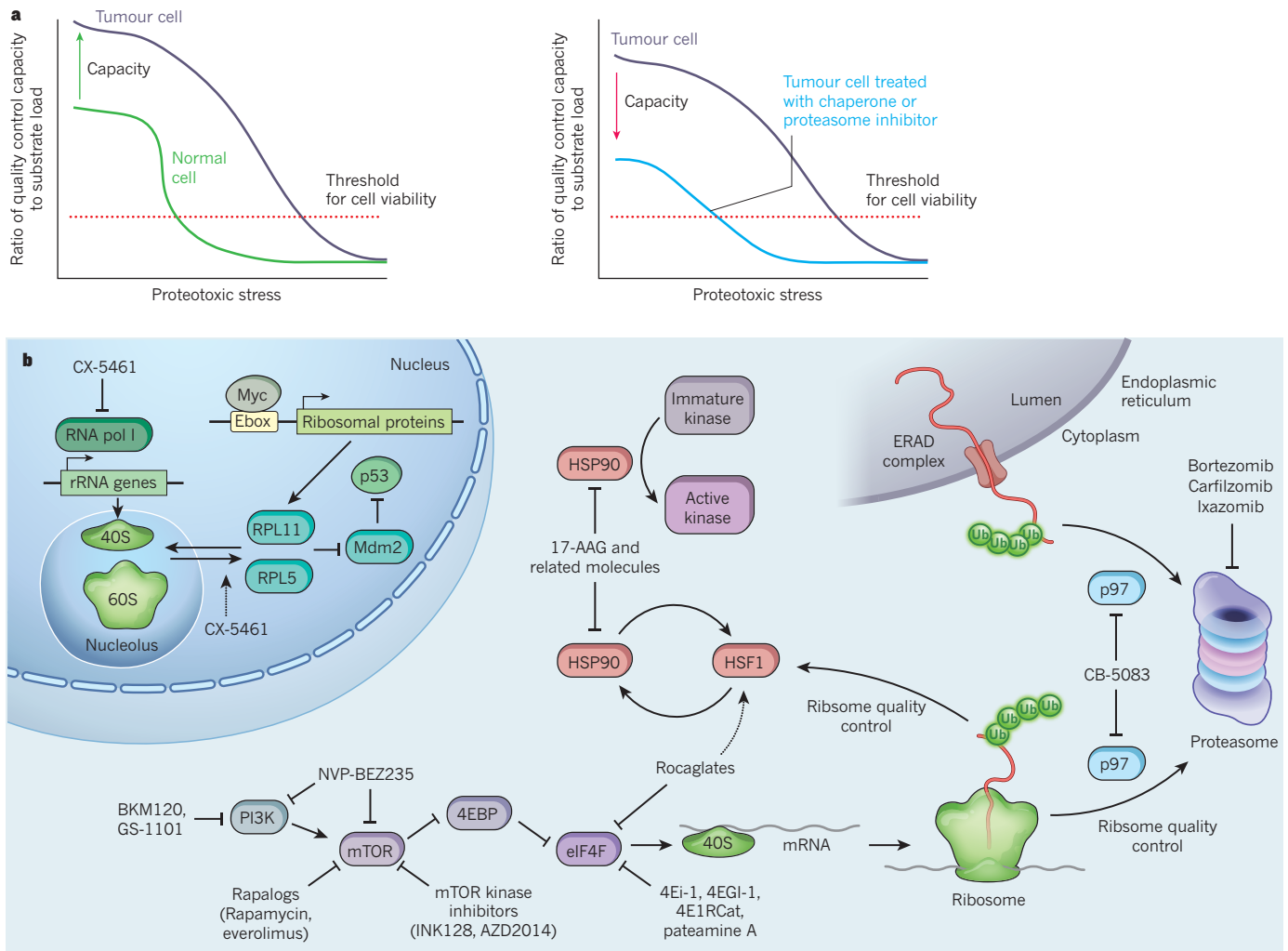


Figure 4 | Therapeutic strategies that target proteome maintenance. **a**, The ability of cells to respond to fluctuations in proteome balance is dictated by the cellular ratio of quality control capacity to substrate abundance. When the capacity to load ratio (y axis) is high, levels of proteotoxic stress (x axis) are low. Tumour cells may adapt to increase their quality control capacity, which enables them to become more resistant to the effects of proteotoxic stress (left). Pharmacological targeting of the pathways that regulate the balance of the proteome with drugs that inhibit chaperones or the proteasome can alter the cell's capacity to load ratio in favour of enhanced cell death at lower levels of proteotoxic stress (right). **b**, A number of pathways that regulate proteome balance have been targeted using small molecules. In the nucleus, inhibitors of RNA pol I, such as CX-5461, block the production of rRNA. This isolated reduction of rRNA leads to an excess of ribosomal proteins RPL11 and RPL5,

recruitment¹¹¹. Evidence to support an increase in the number of autophagosomes and in the expression of autophagic machinery was initially observed in human colon cancer cells (HCT116) that contain supernumerary chromosomes¹¹². However, ubiquitin-binding protein p62 (also known as SQSTM1) — a cargo adaptor that associates with ubiquitylated protein aggregates in HCT116 cells — increases in abundance, which is counter-intuitive if the autophagy flux has increased¹¹². Lipidation of MAP1LC3s and the levels of p62 were also increased in retinal pigment epithelial cells that contain supernumerary chromosomes¹¹³; however, although these MAP1LC3-positive structures were delivered to the lysosome, they were not efficiently degraded. The inability of lysosomes to degrade MAP1LC3s did not reflect the inhibition of lysosomal proteases, but it was accompanied by the activation of a lysosome stress response that involved the transcription factor TFEB¹¹³. It is unclear whether this is a response to the broad ubiquitylation of misfolded proteins. Further studies should help to explain the role

which induces p53 stabilization through inhibition of the ubiquitin ligase Mdm2. In the cytoplasm, translational control mechanisms, protein chaperones such as HSP90 and the AAA-ATPase p97 control a network of interactions that regulate the production and turnover of the defective products of translation at the ribosome. The excess of ribosomal proteins such as RPL11 and RPL5 is also increased in the context of Myc overexpression, which promotes the production of ribosomal proteins. Several small-molecule inhibitors, including the HSP90 inhibitor tanesipimycin (17-AAG), the PI3K and mTOR inhibitor NVP-BEZ235 and various elongation factor 4E (EIF4E) inhibitors (4Ei-1 related molecules and the natural product pateamine A), have been developed to target distinct steps in the pathway. p97 also controls the proteasomal turnover of proteins through the ERAD pathway, which might be important for determining the clinical activity of p97 inhibitor CB-5083. Ub, ubiquitin.

that autophagy has in the catabolism of excess proteins that emerge from supernumerary chromosomes as well as the relationship between autophagy and the ubiquitin system in this context.

Therapeutic targeting of proteome imbalance

At present, there is great interest in determining whether interference in stress-response pathways in cells with elevated proteome imbalance could be used therapeutically, especially given that the threshold for cell viability could be altered (Fig. 4a). Perhaps the most advanced example of using proteome imbalance to treat diseases in people is the application of proteasome inhibitors (such as bortezomib, carfilzomib and ixazomib) in malignant bone-marrow cells^{114,115}. These cells are particularly sensitive to proteasome inhibition because they show a high rate of protein synthesis through the secretory pathway¹¹⁶. This concept has been extended to the p97 (or VCP) AAA-ATPase, which functions as a segregase for the extraction of ubiquitylated proteins from

complexes and membranes, especially in the context of endoplasmic-reticulum-associated degradation (ERAD), and from chromatin¹¹⁷. After retrotranslocation, misfolded and ubiquitylated proteins in the endoplasmic reticulum are extracted and delivered to the proteasome through p97 (ref. 118) (Fig. 4b). ATP-competitive inhibitors of p97 such as CB-5083, which block the delivery of ERAD substrates and other ubiquitylated proteins to the proteasome, have been shown to inhibit the growth of myeloma cells both *in vitro* and in mouse xenograft experiments^{119,120}. But unlike bortezomib, CB-5083 also shows activity towards solid tumours in mice, which generally have more complex karyotypes than do multiple myeloma cells, suggesting that the inhibition of p97 might have an enhanced effect on cells with greater imbalance in the proteome. Interestingly, p97 has roles in the extraction of ubiquitylated proteins from stalled ribosomes⁵⁴ (Fig. 4b). It is possible that some of the effects of p97 inhibition on cell proliferation are a reflection of increased proteotoxic stress that results from a decreased ability to free the ribosome of defective products of translation.

A further target for therapeutics that exploit proteome imbalance has emerged from two independent lines of investigation that point to inhibitors of the heat-shock protein HSP90 as a possible treatment for aneuploid tumours. Two studies in aneuploid model systems^{121,122} led to the finding that HSP90 inhibitors, including tanespimycin (also known as 17-AAG), enhance cell death in cells with aneuploidy compared to non-aneuploid cells. This is consistent with a role for the Hsp90–Cdc37 system in folding and activating protein kinases that are important for proliferation; for example, the toxicity of human SRC kinase expression in yeast is relaxed in disomic cells⁸⁵. This effect is interpreted as there being insufficient levels of the Hsp90–Cdc37 system to support the maturation of kinases in aneuploid cells⁸⁵. The selective inhibition of cells with supernumerary chromosomes by Hsp90 inhibitors has implications for the many clinical trials in progress that are targeting Hsp90 (ref. 123). More broadly, HSF1 might represent an important target for therapeutics. HSF1 promotes homeostasis through the transcription of molecular chaperones, but it is also linked to energy metabolism¹²⁴ (Fig. 2b). The inhibition of translation with small-molecule rocaglates leads to the inactivation of HSF1, a reduction in cellular energy, the loss of chaperone capacity and a selective reduction in the proliferation of cancer cells¹²⁴. Further studies will improve understanding of the mechanistic regulation of HSF1 by alterations in protein synthesis.

Considerable focus has been placed on targeting pathways that promote translation^{22,26} (Fig. 4b), including efforts that have yielded inhibitors of the binding of eIF4E to eIF4F, some of which show efficacy in xenograft models of cancer¹²⁵. Because of the prevalence of Myc activation in cancer, numerous strategies to target Myc function have been implemented⁶², including the inhibition of ribosome biogenesis through RNA polymerase I (ref. 126). The idea that underlies this approach is to inhibit rRNA synthesis without affecting the synthesis of ribosomal proteins, which leads to the continuous production of ribosomal proteins that require degradation. This strategy simultaneously activates the ribosomal surveillance pathway and reduces the capacity for degradation (owing to an excess of orphan ribosomal proteins), and it has proven effective and synergistic with mTOR inhibition in mouse models of Myc-driven B-cell lymphomas¹²⁷. Although the small-molecule-mediated inhibition of RNA polymerase I has not been tested specifically in cells with supernumerary chromosomes, it is intriguing to postulate that a sudden increase in unassembled ribosomal proteins that require quality-control-dependent degradation might have large effects on cells that are already burdened by excess ribosomal protein synthesis. These approaches underscore the general hypothesis that therapeutics aimed at increasing proteotoxic stress in cells with an already overburdened protein-degradation system might represent anti-cancer strategies with broad benefits (Fig. 4a).

Future directions

We now have an extensive picture of how imbalance in the proteome, whether it is generated through the activation of oncogenes or by

BOX 2

Proteome imbalance at a glance

● **Which E3 ubiquitin-ligase systems and mechanisms control the turnover of excess subunits from complexes?** Our understanding of the roles and mechanisms of E3 ligases that are involved in quality control is limited, and the rules for selecting orphan subunits of multimeric complexes for degradation are unclear.

● **What are the mechanisms that underlie spatially distinct quality-control pathways?** Misfolded proteins in yeast are often channelled into three locations: the juxtanuclear quality-control compartment in which active re-folding is promoted; the cytoplasmic insoluble protein deposits compartment for irreconcilably misfolded proteins; or the Q-bodies, which are processing centres for soluble misfolded proteins that are controlled by the Hsp104 disaggregase^{135,136}. Comparatively little is known in mammalian cells about the spatial organization of the synthesis and degradation machinery and whether they are coupled.

● **Which mechanisms underlie the capacity of the cell for degradation?** The occupancy of the proteasome is estimated to be about 20%, based on tomography in neurons³⁸, but it is unclear whether this reflects the full capacity of the cell's proteolytic machinery. The finding that deletion of the deubiquitinase Ubp6 in yeast (or of USP14 in mammalian cells) can increase flux through the proteasome suggests that it is possible to increase proteasomal activity, although perhaps at the expense of specificity.

● **What is the role of autophagy in controlling proteome imbalance?** Links between autophagy and proteotoxic stress in the context of chromosome imbalance are limited and the circumstances under which autophagy is activated or inhibited are unclear. The extent to which proteotoxic stress produces protein aggregates that require autophagy for degradation is unknown. The basis for the apparent inhibition of protein degradation within lysosomes in the context of chromosome imbalance¹¹³ is also unknown.

● **What are the molecular determinants that specify the ubiquitylation and turnover of defective products of translation?** Ribosomes that lack the ubiquitin ligase Ltn1 are unable to degrade misfolded nascent chains efficiently; this leads to the accumulation of nascent proteins with C-terminal alanine and threonine extensions, which points to a central role for Ltn1 in the removal of defective nascent chains. However, Ltn1 is present at levels that are much lower than those of the ribosome. It will be important to understand the dynamics of Ltn1 and whether there are mechanisms that can compensate for its loss or operate in parallel in the context of particular types of translational errors.

● **What contributions do membrane proteins make to proteome imbalance?** The answer to this question will probably require the development of improved methods to quantify the turnover of the pools of individual proteins that are contained within vesicular structures as well as the turnover of membrane proteins themselves.

alterations in gene dosage, limits the capacity of systems that chaperone or degrade proteins. Studies of gene dosage, in particular, have revealed the existence of widespread dosage compensation that is mediated through the ubiquitin–proteasome system, and this finding is beginning to shed light on how the stoichiometry of proteins is established within the cell. Numerous questions have emerged (Box 2), such as how and from where do cells orchestrate the degradation of excess subunits from protein complexes, and whether there are differences in the capacity of distinct types of cells to handle proteome imbalance. In this regard,

haematopoietic stem cells have lower rates of protein synthesis *in vivo* in comparison to those of differentiated lineages, and they are sensitive to genetic perturbations that lead to either increases or decreases in protein synthesis^{128,129}. Attempts to quantify proteomes have focused on identification of the relative abundances of proteins within ensembles of cells¹³⁰. However, such studies do not address features of the proteome that are crucial for understanding how balance is achieved. In particular, studies that are limited to measuring the total abundance of proteins do not address the complexity that individual proteins exhibit through various free and complexed forms that might have broadly different stabilities. Understanding turnover rates for both orphan subunits and their assemblies on a global scale will require innovative methods that rely on the integration of quantitative proteomics and imaging. Moreover, a present limitation is that most studies that analyse protein abundance do not reach the depth that is required to fully address the status of low-abundance proteins (Box 1), and it is unclear whether the behaviour of abundant proteins accurately models that of proteins that are near the limits of detection. The reliance on using bulk measurements to understand most assembly and turnover pathways limits our ability to appreciate fully the underlying regulatory systems. In the future, methods based on analyses of single cells will help to unravel the complex interplay between protein quality control and proteome imbalance.

Note added in proof: Recent work has demonstrated a conserved pathway in which mTORC1 negatively regulates the production of assembly factors and chaperones called ribosome-associated complexes (RACs) for the proteasome (A. Rousseau and A. Bertolotti. An evolutionarily conserved pathway controls proteasome homeostasis *Nature* **536**, 184–189; 2016). mTORC1 inhibition leads to the phosphorylation of MAP kinase family members, which promotes RAC expression. This suggests that proteasome activity is controlled by a highly integrated network and provides a therapeutic target. ■

Received 28 January; accepted 29 July 2016.

- Nagaraj, N. *et al.* Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* **7**, 548 (2011).
- Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
- Wiśniewski, J. R., Hein, M. Y., Cox, J. & Mann, M. A “proteomic ruler” for protein copy number and concentration estimation without spike-in standards. *Mol. Cell. Proteomics* **13**, 3497–3506 (2014).
- Jünger, M. A. & Aebersold, R. Mass spectrometry-driven phosphoproteomics: patterning the systems biology mosaic. *Wiley Interdiscip. Rev. Dev. Biol.* **3**, 83–112 (2014).
- Hornbeck, P. V. *et al.* PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43**, D512–D520 (2015).
- Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genet.* **40**, 1413–1415 (2008).
- Buszczak, M., Signer, R. A. & Morrison, S. J. Cellular differences in protein synthesis regulate tissue homeostasis. *Cell* **159**, 242–251 (2014).
- Jovanovic, M. *et al.* Dynamic profiling of the protein life cycle in response to pathogens. *Science* **347**, 1259038 (2015).
- Liu, Y., Beyer, A. & Aebersold, R. On the dependency of cellular protein levels on mRNA abundance. *Cell* **165**, 535–550 (2016).
- The 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
- Beroukhim, R. *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905 (2010).
- Garraway, L. A. & Lander, E. S. Lessons from the cancer genome. *Cell* **153**, 17–37 (2013).
- Zhang, B. *et al.* Proteogenomic characterization of human colon and rectal cancer. *Nature* **513**, 382–387 (2014).
- Ingolia, N. T. Ribosome profiling: new views of translation, from single codons to genome scale. *Nature Rev. Genet.* **15**, 205–213 (2014).
- Kim, M. S. *et al.* A draft map of the human proteome. *Nature* **509**, 575–581 (2014).
- Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Rev. Genet.* **13**, 227–232 (2012).
- Wang, M., Herrmann, C. J., Simonovic, M., Szklarczyk, D. & von Mering, C. Version 4.0 of PaxDb: protein abundance data, integrated across model organisms, tissues, and cell-lines. *Proteomics* **15**, 3163–3168 (2015).
- Wilhelm, M. *et al.* Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582–587 (2014).
- Dever, T. E. & Green, R. The elongation, termination, and recycling phases of translation in eukaryotes. *Cold Spring Harb. Perspect. Biol.* **4**, a013706 (2012).
- Rodrigo-Brenni, M. C. & Hegde, R. S. Design principles of protein biosynthesis-coupled quality control. *Dev. Cell* **23**, 896–907 (2012).
- Sonenberg, N. & Hinnebusch, A. G. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136**, 731–745 (2009).
- Bhat, M. *et al.* Targeting the translation machinery in cancer. *Nature Rev. Drug Discov.* **14**, 261–278 (2015).
- Deshai, R. J. Proteotoxic crisis, the ubiquitin-proteasome system, and cancer therapy. *BMC Biol.* **12**, 94 (2014).
- Laplanche, M. & Sabatini, D. M. mTOR signaling in growth control and disease. *Cell* **149**, 274–293 (2012).
- Hsieh, A. L., Walton, Z. E., Altman, B. J., Stine, Z. E. & Dang, C. V. MYC and metabolism on the path to cancer. *Semin. Cell Dev. Biol.* **43**, 11–21 (2015).
- Ruggero, D. Translational control in cancer etiology. *Cold Spring Harb. Perspect. Biol.* **5**, (2013).
- Wong, K. K., Engelman, J. A. & Cantley, L. C. Targeting the PI3K signaling pathway in cancer. *Curr. Opin. Genet. Dev.* **20**, 87–90 (2010).
- Basan, M. *et al.* Inflating bacterial cells by increased protein synthesis. *Mol. Syst. Biol.* **11**, 836 (2015).
- Scott, M., Klumpp, S., Mateescu, E. M. & Hwa, T. Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Mol. Syst. Biol.* **10**, 747 (2014).
- Drummond, D. A. & Wilke, C. O. The evolutionary consequences of erroneous protein synthesis. *Nature Rev. Genet.* **10**, 715–724 (2009).
- Zaher, H. S. & Green, R. Fidelity at the molecular level: lessons from protein synthesis. *Cell* **136**, 746–762 (2009).
- Doherty, M. K., Hammond, D. E., Clague, M. J., Gaskell, S. J. & Beynon, R. J. Turnover of the human proteome: determination of protein intracellular stability by dynamic SILAC. *J. Proteomics Res.* **8**, 104–112 (2009).
- Boisvert, F.-M. *et al.* A quantitative spatial proteomics analysis of proteome turnover in human cells. *Mol. Cell. Proteomics* **11**, M111.011429 (2012).
- Duttler, S., Pechmann, S. & Frydman, J. Principles of cotranslational ubiquitination and quality control at the ribosome. *Mol. Cell* **50**, 379–393 (2013).
- Wang, F., Durfee, L. A. & Huibregtse, J. M. A cotranslational ubiquitination pathway for quality control of misfolded proteins. *Mol. Cell* **50**, 368–378 (2013).
- Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789–802 (2011).
- Peth, A., Nathan, J. A. & Goldberg, A. L. The ATP costs and time required to degrade ubiquitinated proteins by the 26 S proteasome. *J. Biol. Chem.* **288**, 29215–29222 (2013).
- Asano, S. *et al.* A molecular census of 26S proteasomes in intact neurons. *Science* **347**, 439–442 (2015).
- Gendron, J. M. *et al.* Using the ubiquitin-modified proteome to monitor distinct and spatially restricted protein homeostasis dysfunction. *Mol. Cell. Proteomics* **15**, 2576–2593 (2016).
- Ishimura, R. *et al.* Ribosome stalling induced by mutation of a CNS-specific tRNA causes neurodegeneration. *Science* **345**, 455–459 (2014).
- This paper and ref. 41 identify mutations in mice that increase miscoding during translation and result in neurodegenerative phenotypes.**
- Lee, J. W. *et al.* Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. *Nature* **443**, 50–55 (2006).
- Chu, J. *et al.* A mouse forward genetics screen identifies LISTERIN as an E3 ubiquitin ligase involved in neurodegeneration. *Proc. Natl Acad. Sci. USA* **106**, 2097–2103 (2009).
- Triman, K. L. Mutational analysis of the ribosome. *Adv. Genet.* **58**, 89–119 (2007).
- Zaher, H. S. & Green, R. Hyperaccurate and error-prone ribosomes exploit distinct mechanisms during tRNA selection. *Mol. Cell* **39**, 110–120 (2010).
- Paredes, J. A. *et al.* Low level genome mistranslations deregulate the transcriptome and translate and generate proteotoxic stress in yeast. *BMC Biol.* **10**, 55 (2012).
- Kalapis, D. *et al.* Evolution of robustness to protein mistranslation by accelerated protein turnover. *PLoS Biol.* **13**, e1002291 (2015).
- This paper provides evidence of adaptation to proteotoxic stress through large chromosomal alterations and the genetic modification of protein-degradation systems.**
- Bengtson, M. H. & Joazeiro, C. A. Role of a ribosome-associated E3 ubiquitin ligase in protein quality control. *Nature* **467**, 470–473 (2010).
- Defenouillère, Q. *et al.* Cdc48-associated complex bound to 60S particles is required for the clearance of aberrant translation products. *Proc. Natl Acad. Sci. USA* **110**, 5046–5051 (2013).
- Pisareva, V. P., Skabkin, M. A., Hellen, C. U., Pestova, T. V. & Pisarev, A. V. Dissociation by Pelota, Hbs1 and ABCE1 of mammalian vacant 80S ribosomes and stalled elongation complexes. *EMBO J.* **30**, 1804–1817 (2011).
- Shao, S. & Hegde, R. S. Reconstitution of a minimal ribosome-associated ubiquitination pathway with purified factors. *Mol. Cell* **55**, 880–890 (2014).
- This paper and ref. 50 reconstitute Ltn1-mediated RQC and establish the order of operations.**
- Shao, S., von der Malsburg, K. & Hegde, R. S. Listerin-dependent nascent protein ubiquitination relies on ribosome subunit dissociation. *Mol. Cell* **50**, 637–648 (2013).
- Shoemaker, C. J., Eyler, D. E. & Green, R. Dom34:Hbs1 promotes subunit dissociation and peptidyl-tRNA drop-off to initiate no-go decay. *Science* **330**, 369–372 (2010).

53. Verma, R., Oania, R. S., Kolawa, N. J. & Deshaies, R. J. Cdc48/p97 promotes degradation of aberrant nascent polypeptides bound to the ribosome. *eLife* **2**, e00308 (2013).
54. Brandman, O. & Hegde, R. S. Ribosome-associated protein quality control. *Nature Struct. Mol. Biol.* **23**, 7–15 (2016).
55. Shen, P. S. *et al.* Rqc2p and 60S ribosomal subunits mediate mRNA-independent elongation of nascent chains. *Science* **347**, 75–78 (2015). **Refs 55, 57 and 59 identify the mechanism of RQC-mediated CAT-tail formation and demonstrate that defective tail formation suppresses the protein aggregation that results from RQC events.**
56. Brandman, O. *et al.* A ribosome-bound quality control complex triggers degradation of nascent peptides and signals translation stress. *Cell* **151**, 1042–1054 (2012).
57. Choe, Y. J. *et al.* Failure of RQC machinery causes protein aggregation and proteotoxic stress. *Nature* **531**, 191–195 (2016).
58. Yang, J., Hao, X., Cao, X., Liu, B. & Nystrom, T. Spatial sequestration and detoxification of Huntingtin by the ribosome quality control complex. *eLife* **5**, e11792 (2016).
59. Yonashiro, R. *et al.* The Rqc2/Tae2 subunit of the ribosome-associated quality control (RQC) complex marks ribosome-stalled nascent polypeptide chains for aggregation. *eLife* **5**, e11794 (2016).
60. Higgins, R. *et al.* The unfolded protein response triggers site-specific regulatory ubiquitylation of 40S ribosomal proteins. *Mol. Cell* **59**, 35–49 (2015).
61. Loayza-Puch, F. *et al.* Tumour-specific proline vulnerability uncovered by differential ribosome codon reading. *Nature* **530**, 490–494 (2016).
62. Poortinga, G., Quinn, L. M. & Hannan, R. D. Targeting RNA polymerase I to treat MYC-driven cancer. *Oncogene* **34**, 403–412 (2015).
63. Barna, M. *et al.* Suppression of Myc oncogenic activity by ribosomal protein haploinsufficiency. *Nature* **456**, 971–975 (2008). **This paper demonstrates that a Myc-driven model of lymphoma can be suppressed by the haploinsufficiency of specific ribosomal genes.**
64. Conn, C. S. & Qian, S. B. Nutrient signaling in protein homeostasis: an increase in quantity at the expense of quality. *Sci. Signal.* **6**, ra24 (2013).
65. Mitsiades, N. *et al.* Molecular sequelae of proteasome inhibition in human multiple myeloma cells. *Proc. Natl Acad. Sci. USA* **99**, 14374–14379 (2002).
66. Meiners, S. *et al.* Inhibition of proteasome activity induces concerted expression of proteasome genes and *de novo* formation of mammalian proteasomes. *J. Biol. Chem.* **278**, 21517–21525 (2003).
67. Radhakrishnan, S. K. *et al.* Transcription factor Nrf1 mediates the proteasome recovery pathway after proteasome inhibition in mammalian cells. *Mol. Cell* **38**, 17–28 (2010).
68. Ruggero, D. *et al.* The translation factor eIF-4E promotes tumor formation and cooperates with c-Myc in lymphomagenesis. *Nature Med.* **10**, 484–486 (2004).
69. Avdulov, S. *et al.* Activation of translation complex eIF4F is essential for the genesis and maintenance of the malignant phenotype in human mammary epithelial cells. *Cancer Cell* **5**, 553–563 (2004).
70. Lazaris-Karatzas, A., Montine, K. S. & Sonenberg, N. Malignant transformation by a eukaryotic initiation factor subunit that binds to mRNA 5' cap. *Nature* **345**, 544–547 (1990).
71. Hsieh, A. C. *et al.* Genetic dissection of the oncogenic mTOR pathway reveals druggable addiction to translational control via 4EBP-eIF4E. *Cancer Cell* **17**, 249–261 (2010).
72. Hsieh, A. C. *et al.* The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* **485**, 55–61 (2012).
73. Truitt, M. L. *et al.* Differential requirements for eIF4E dose in normal development and cancer. *Cell* **162**, 59–71 (2015). **Refs 73 and 74 establish that tumorigenesis can be suppressed *in vivo* by genetically limiting eIF4E.**
74. Hsieh, A. C. *et al.* Cell type-specific abundance of 4EBP1 primes prostate cancer sensitivity or resistance to PI3K pathway inhibitors. *Sci. Signal.* **8**, ra116 (2015).
75. Katz, W., Weinstein, B. & Solomon, F. Regulation of tubulin levels and microtubule assembly in *Saccharomyces cerevisiae*: consequences of altered tubulin gene copy number. *Mol. Cell. Biol.* **10**, 5286–5294 (1990).
76. Weinstein, B. & Solomon, F. Phenotypic consequences of tubulin overproduction in *Saccharomyces cerevisiae*: differences between alpha-tubulin and beta-tubulin. *Mol. Cell. Biol.* **10**, 5295–5304 (1990).
77. Li, G. W., Burkhardt, D., Gross, C. & Weissman, J. S. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **157**, 624–635 (2014).
78. Torres, E. M. *et al.* Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* **317**, 916–924 (2007). **This study provided the first systematic analysis of both the effects of aneuploidy of intact chromosomes in yeast and the effects of an increased chromosome copy number on cell fitness.**
79. Sheltzer, J. M. *et al.* Aneuploidy drives genomic instability in yeast. *Science* **333**, 1026–1030 (2011).
80. Thornburn, R. R. *et al.* Aneuploid yeast strains exhibit defects in cell growth and passage through START. *Mol. Biol. Cell* **24**, 1274–1289 (2013).
81. Blank, H. M., Sheltzer, J. M., Meehl, C. M. & Amon, A. Mitotic entry in the presence of DNA damage is a widespread property of aneuploidy in yeast. *Mol. Biol. Cell* **26**, 1440–1451 (2015).
82. Dephousse, N. *et al.* Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast. *eLife* **3**, e03023 (2014). **This paper provides a comprehensive proteomic and transcriptional analysis of yeast disomic strains, demonstrating that a subset of genes on disomes undergo dosage compensation through active mechanisms of turnover.**
83. Tang, Y. C. & Amon, A. Gene copy-number alterations: a cost-benefit analysis. *Cell* **152**, 394–405 (2013).
84. Bonney, M. E., Moriya, H. & Amon, A. Aneuploid proliferation defects in yeast are not driven by copy number changes of a few dosage-sensitive genes. *Genes Dev.* **29**, 898–903 (2015).
85. Oromendia, A. B., Dodgson, S. E. & Amon, A. Aneuploidy causes proteotoxic stress in yeast. *Genes Dev.* **26**, 2696–2708 (2012). **This paper reveals that aneuploidy in yeast generates proteotoxic stress, which affects the ability of yeast to properly fold proteins that are already prone to misfolding.**
86. Sheltzer, J. M., Torres, E. M., Dunham, M. J. & Amon, A. Transcriptional consequences of aneuploidy. *Proc. Natl Acad. Sci. USA* **109**, 12644–12649 (2012).
87. Gasch, A. P. *et al.* Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11**, 4241–4257 (2000).
88. Crosas, B. *et al.* Ubiquitin chains are remodeled at the proteasome by opposing ubiquitin ligase and deubiquitinating activities. *Cell* **127**, 1401–1413 (2006).
89. Torres, E. M. *et al.* Identification of aneuploidy-tolerating mutations. *Cell* **143**, 71–83 (2010).
90. Tasaki, T., Sriram, S. M., Park, K. S. & Kwon, Y. T. The N-end rule pathway. *Annu. Rev. Biochem.* **81**, 261–289 (2012).
91. Hwang, C. S., Shemorry, A. & Varshavsky, A. N-terminal acetylation of cellular proteins creates specific degradation signals. *Science* **327**, 973–977 (2010).
92. Shemorry, A., Hwang, C. S. & Varshavsky, A. Control of protein quality and stoichiometries by N-terminal acetylation and the N-end rule pathway. *Mol. Cell* **50**, 540–551 (2013). **This paper provides evidence for N-acetylation in the turnover of conserved oligomeric Golgi and anaphase-promoting complex proteins that are not properly assembled into their associated complexes.**
93. Arnesen, T. *et al.* Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *Proc. Natl Acad. Sci. USA* **106**, 8157–8162 (2009).
94. Van Damme, P. *et al.* NatF contributes to an evolutionary shift in protein N-terminal acetylation and is important for normal chromosome segregation. *PLoS Genet.* **7**, e1002169 (2011).
95. Scott, D. C., Monda, J. K., Bennett, E. J., Harper, J. W. & Schulman, B. A. N-terminal acetylation acts as an avidity enhancer within an interconnected multiprotein complex. *Science* **334**, 674–678 (2011).
96. McClellan, A. J., Scott, M. D. & Frydman, J. Folding and quality control of the VHL tumor suppressor proceed through distinct chaperone pathways. *Cell* **121**, 739–748 (2005).
97. Scazzari, M., Amm, I. & Wolf, D. H. Quality control of a cytoplasmic protein complex: chaperone motors and the ubiquitin-proteasome system govern the fate of orphan fatty acid synthase subunit Fas2 of yeast. *J. Biol. Chem.* **290**, 4677–4687 (2015).
98. Schüller, H. J., Förtsch, B., Rautenstrauss, B., Wolf, D. H. & Schweizer, E. Differential proteolytic sensitivity of yeast fatty acid synthetase subunits α and β contributing to a balanced ratio of both fatty acid synthetase components. *Eur. J. Biochem.* **203**, 607–614 (1992).
99. Kim, W. *et al.* Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol. Cell* **44**, 325–340 (2011). **This study provided the first large-scale analysis of the ubiquitin-modified proteome.**
100. Warner, J. R. In the absence of ribosomal RNA synthesis, the ribosomal proteins of HeLa cells are synthesized normally and degraded rapidly. *J. Mol. Biol.* **115**, 315–333 (1977).
101. Abovich, N., Gritz, L., Tung, L. & Rosbash, M. Effect of *RP51* gene dosage alterations on ribosome synthesis in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **5**, 3429–3435 (1985).
102. Sung, M. K., Reitsma, J. M., Sweredoski, M. J., Hess, S. & Deshaies, R. J. Ribosomal proteins produced in excess are degraded by the ubiquitin-proteasome system. *Mol. Biol. Cell* <http://dx.doi.org/10.1091/mbc.E16-05-0290> (2016).
103. Bursac, S. *et al.* Mutual protection of ribosomal proteins L5 and L11 from degradation is essential for p53 activation upon ribosomal biogenesis stress. *Proc. Natl Acad. Sci. USA* **109**, 20467–20472 (2012).
104. Deisenroth, C. & Zhang, Y. Ribosome biogenesis surveillance: probing the ribosomal protein-Mdm2-p53 pathway. *Oncogene* **29**, 4253–4260 (2010).
105. Hessa, T. *et al.* Protein targeting and degradation are coupled for elimination of mislocalized proteins. *Nature* **475**, 394–397 (2011).
106. Shao, S. & Hegde, R. S. Target selection during protein quality control. *Trends Biochem. Sci.* **41**, 124–137 (2016).
107. Wang, Q. *et al.* A ubiquitin ligase-associated chaperone holdase maintains polypeptides in soluble states for proteasome degradation. *Mol. Cell* **42**, 758–770 (2011).
108. Rodrigo-Brenni, M. C., Gutierrez, E. & Hegde, R. S. Cytosolic quality control of mislocalized proteins requires RNF126 recruitment to Bag6. *Mol. Cell* **55**, 227–237 (2014).
109. Itakura, E. *et al.* Ubiquitins chaperone and triage mitochondrial membrane proteins for degradation. *Mol. Cell* **63**, 21–33 (2016).
110. Wrobel, L. *et al.* Mistargeted mitochondrial proteins activate a proteostatic response in the cytosol. *Nature* **524**, 485–488 (2015).
111. Mancias, J. D. & Kimmelman, A. C. Mechanisms of selective autophagy in normal physiology and cancer. *J. Mol. Biol.* **428**, 1659–1680 (2016).
112. Stinglee, S. *et al.* Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. *Mol. Syst. Biol.* **8**, 608 (2012).

113. Santaguida, S., Vasile, E., White, E. & Amon, A. Aneuploidy-induced cellular stresses limit autophagic degradation. *Genes Dev.* **29**, 2010–2021 (2015).
This paper demonstrates that autophagy is induced rapidly after chromosome mis-segregation in mammalian cells and that undigested proteins accumulate in lysosomes after delivery by autophagosomes.
114. Dou, Q. P. & Zonder, J. A. Overview of proteasome inhibitor-based anti-cancer therapies: perspective on bortezomib and second generation proteasome inhibitors versus future generation inhibitors of ubiquitin-proteasome system. *Curr. Cancer Drug Targets* **14**, 517–536 (2014).
115. Moreau, P. *et al.* Oral ixazomib, lenalidomide, and dexamethasone for multiple myeloma. *N. Engl. J. Med.* **374**, 1621–1634 (2016).
116. San Miguel, J. Multiple myeloma: a model for scientific and clinical progress. *Hematology Am. Soc. Hematol. Educ. Program* **2014**, 1–7 (2014).
117. Meyer, H., Bug, M. & Bremer, S. Emerging functions of the VCP/p97 AAA-ATPase in the ubiquitin system. *Nature Cell Biol.* **14**, 117–123 (2012).
118. Christianson, J. C. & Ye, Y. Cleaning up in the endoplasmic reticulum: ubiquitin in charge. *Nature Struct. Mol. Biol.* **21**, 325–335 (2014).
119. Zhou, H. J. *et al.* Discovery of a first-in-class, potent, selective, and orally bioavailable inhibitor of the p97 AAA ATPase (CB-5083). *J. Med. Chem.* **58**, 9480–9497 (2015).
120. Anderson, D. J. *et al.* Targeting the AAA ATPase p97 as an approach to treat cancer through disruption of protein homeostasis. *Cancer Cell* **28**, 653–665 (2015).
This paper identified inhibitors of the AAA-ATPase p97 and demonstrated their activity against cancer cells both *in vitro* and in xenograft experiments.
121. Donnelly, N., Passerini, V., Durrbaum, M., Stingle, S. & Storchova, Z. HSF1 deficiency and impaired HSP90-dependent protein folding are hallmarks of aneuploid human cells. *EMBO J.* **33**, 2374–2387 (2014).
122. Tang, Y. C., Williams, B. R., Siegel, J. J. & Amon, A. Identification of aneuploidy-selective antiproliferation compounds. *Cell* **144**, 499–512 (2011).
123. Jhaveri, K., Taldone, T., Modi, S. & Chiosis, G. Advances in the clinical development of heat shock protein 90 (Hsp90) inhibitors in cancers. *Biochim. Biophys. Acta* **1823**, 742–755 (2012).
124. Santagata, S. *et al.* Tight coordination of protein translation and HSF1 activation supports the anabolic malignant state. *Science* **341**, 1238303 (2013).
125. Pelletier, J., Graff, J., Ruggero, D. & Sonenberg, N. Targeting the eIF4F translation initiation complex: a critical nexus for cancer development. *Cancer Res.* **75**, 250–263 (2015).
126. Drygin, D. *et al.* Targeting RNA polymerase I with an oral small molecule CX-5461 inhibits ribosomal RNA synthesis and solid tumor growth. *Cancer Res.* **71**, 1418–1430 (2011).
127. Devlin, J. R. *et al.* Combination therapy targeting ribosome biogenesis and mRNA translation synergistically extends survival in MYC-driven lymphoma. *Cancer Discov.* **6**, 59–70 (2016).
128. Vilchez, D. *et al.* Increased proteasome activity in human embryonic stem cells is regulated by PSMD11. *Nature* **489**, 304–308 (2012).
Refs 128 and 129 establish that some populations of stem cells require lower rates of protein synthesis and show elevated capacities for protein degradation.
129. Signer, R. A., Magee, J. A., Salic, A. & Morrison, S. J. Haematopoietic stem cells require a highly regulated protein synthesis rate. *Nature* **509**, 49–54 (2014).
130. Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–2526 (2014).
131. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
132. Vuckovic, D., Dagley, L. F., Purcell, A. W. & Emili, A. Membrane proteomics by high performance liquid chromatography-tandem mass spectrometry: analytical approaches and challenges. *Proteomics* **13**, 404–423 (2013).
133. Marguerat, S. *et al.* Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. *Cell* **151**, 671–683 (2012).
134. Huttlin, E. L. *et al.* The BioPlex network: a systematic exploration of the human interactome. *Cell* **162**, 425–440 (2015).
135. Kaganovich, D., Kopito, R. & Frydman, J. Misfolded proteins partition between two distinct quality control compartments. *Nature* **454**, 1088–1095 (2008).
136. Escusa-Toret, S., Vonk, W. I. & Frydman, J. Spatial sequestration of misfolded proteins by a dynamic chaperone pathway enhances cellular fitness during stress. *Nature Cell Biol.* **15**, 1231–1243 (2013).

Acknowledgements J.W.H. is supported by grants from the US National Institutes of Health (AG011085, R37NS083524 and GM095567) and by grants from Biogen, Inc. E.J.B. is supported by grants from the National Institutes of Health (1DP2GM119132 and 2P50GM085764) and by a New Scholar in Aging award from The Ellison Medical Foundation (AG-NS-0902-12). We thank A. Amon at the Massachusetts Institute of Technology for providing comments and insight during the preparation of this manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this paper at go.nature.com/2c4qrb0. Correspondence should be addressed to J.W.H. (wade_harper@hms.harvard.edu) and E.J.B. (e1bennett@ucsd.edu).

Unravelling biological macromolecules with cryo-electron microscopy

Rafael Fernandez-Leiro^{1*} & Sjors H. W. Scheres^{1*}

Knowledge of the three-dimensional structures of proteins and other biological macromolecules often aids understanding of how they perform complicated tasks in the cell. Because many such tasks involve the cleavage or formation of chemical bonds, structural characterization at the atomic level is most useful. Developments in the electron microscopy of frozen hydrated samples (cryo-electron microscopy) are providing unprecedented opportunities for the structural characterization of biological macromolecules. This is resulting in a wave of information about processes in the cell that were impossible to characterize with existing techniques in structural biology.

Biological macromolecules adopt complicated three-dimensional (3D) structures that are crucial to their function. Some macromolecules, such as enzymes, act alone to provide chemical environments that favour the catalysis of specific chemical reactions. Other macromolecules form larger complexes with protein partners, nucleic acids, lipids or sugar molecules. Many such complexes perform their functions through the relative movements of individual parts, in a way that resembles how man-made machines work¹. A fundamental goal of modern biology is to understand how these complicated structures perform their tasks.

Macromolecular complexes are too small to be seen with visible light. Photons with wavelengths that are short enough to visualize details at the atomic level are found in the X-ray region of the electromagnetic spectrum. X-rays interact weakly with biological matter, which makes it difficult to use them to study individual protein complexes. But when many copies of the same protein are arranged into a 3D crystal, information about the atomic structure of the protein can be obtained through X-ray diffraction experiments. This technique, known as X-ray crystallography, has been the most important tool in structural biology for more than two decades (Fig. 1). Another technique that is used to characterize the structures of proteins is nuclear magnetic resonance (NMR) spectroscopy, which measures distance-dependent interactions between atoms. NMR can be used to infer the structure of relatively small proteins, and it provides unique information about the dynamics of proteins and their interactions with other molecules.

Electrons can also be used to look at protein structures. Proteins scatter electrons about ten-thousand times more strongly than they do X-rays, and electrons can be accelerated by electric fields of several hundreds of thousands of volts to wavelengths that are much shorter than the distances between the atoms in protein structures. Moreover, the electric charge of electrons makes it relatively easy to focus them with electromagnetic lenses. Microscopes can therefore be built that use electrons to make images with atomic-level detail. The contributions of electron microscopy to structural biology have been modest in comparison with X-ray crystallography and NMR, but present trends indicate that this is changing (Fig. 1). In 2016, the one-thousandth atomic structure derived from electron microscopy images² was entered into the Protein Data Bank (PDB), the main repository for protein structures. In this Review, we describe how images from an electron microscope can be used to

study the structures of proteins and how rapid progress in structure determination through electron microscopy in the past few years has been heralded as the start of a revolution in structural biology³. We also highlight how the unique characteristics of this technique have already changed structural biology and identify opportunities through which electron microscopy will continue to transform understanding of how macromolecules perform intricate tasks in the cell.

Cryo-electron microscopy

Because electrons are scattered by molecules in the air, electron microscopes must be operated in a high vacuum. This poses a problem when studying biological samples, most of which occur naturally in an aqueous environment. Biological structures are also sensitive to radiation damage. For each electron that contributes to the formation of an image, there will be three electrons that deposit energy in the sample. This energy causes the cleavage of chemical bonds and ultimately destroys the structures of interest. By keeping samples at cryogenic temperatures, cryo-electron microscopy (cryo-EM) enables their preservation in a high vacuum and provides them with some protection against the effects of radiation damage⁴.

The placement of a cryo-EM sample inside an electron microscope is shown in Box 1. To prepare the sample, a few microlitres of a purified protein solution is applied to a metal (usually copper) grid, on top of which lies a thin film of amorphous carbon that contains holes. After any excess liquid is blotted with filter paper, the grid is plunged into liquid ethane⁵. Ideally, this results in the formation of a thin layer of non-crystalline or vitreous ice, in which copies of the protein are deposited in a range of orientations. Images that are captured through the holes in the carbon film contain two-dimensional (2D) projections of individual protein complexes, which are called particles. Projections of particles in various orientations provide complementary information about the underlying 3D object. Numerous 2D projections can therefore be combined into a single 3D reconstruction, provided that their relative orientations are known. Unfortunately, this information is lost in the experiment because the individual particles tumble randomly in solution before the sample is vitrified.

The relative orientations of individual particles can still be determined a posteriori by processing the 2D projection images using a computer⁶, a process known as single-particle analysis. Images

¹MRC Laboratory of Molecular Biology, Cambridge Biomedical Campus, Cambridge CB2 0QH, UK. *These authors contributed equally to this work.

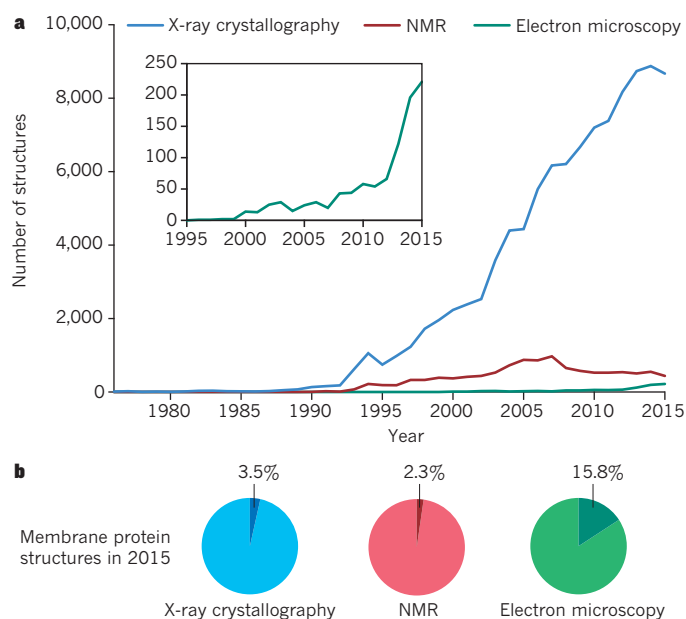


Figure 1 | Growth in structural biology over the past 40 years. **a**, The number of structures recorded in the PDB, as determined by the techniques of X-ray crystallography (blue), NMR (red) and electron microscopy (green) between 1975 and 2015. The number of structures that were determined by electron microscopy between 1995 and 2015 is also shown (inset), which highlights the recent growth in structure determination using this method. **b**, The percentage of membrane protein structures discovered in 2015 using each of the three techniques.

collected with single-particle cryo-EM are of low contrast because the proteins scatter electrons only about 30% more than does the surrounding ice. This would not be a problem if many electrons could be used to determine small differences in image intensities. However, to reduce the effects of radiation damage, the number of electrons used must be carefully limited, resulting in extremely noisy images. Such noise impedes the accurate assignment of orientation to the particles, which becomes the main bottleneck.

Progress in cryo-EM

Larger protein complexes and complexes that contain oligonucleotides give rise to images with higher signal-to-noise ratios. The presence of internal symmetry in a complex also helps to improve the resolution. This explains why the structures of ribosomes and icosahedral virus capsids have long been at the forefront of the cryo-EM field. By 2008, the structures of several viruses had been solved to near-atomic resolution^{7–9}. Ribosome structures could also be calculated to a resolution of 6 Å (refs 10–12). These successes built on important developments in instrumentation, experimental procedures and image processing in the preceding 40 years^{13–15}. And in the past 5 years, a number of developments have further changed the scope of cryo-EM-based structure determination.

By 2010, the quality of available cryo-EM protein structures did not support theoretical considerations about radiation damage, which predicted that structure determination should be possible to atomic resolution for protein complexes with molecular weights as low as 100 kDa (ref. 16). The difference could be explained in part by the inefficient detection of electrons. Images were originally recorded on photographic film, but in the early 2000s, the development of cameras containing a digital charge-coupled device (CCD)¹⁷ opened the path to higher throughput and automated data acquisition¹⁸. Photographic film detects only about one-third of the incoming electrons. CCDs are even less efficient, however, because an extra conversion from electrons to photons leads to an overall detection rate of less than one-fifth of the incoming electrons¹⁹. This affects

cryo-EM-based structure determination exactly at its bottleneck: the low number of electrons that is used to limit radiation damage results in images that are too noisy to determine reliably the orientation of particles.

Three companies had produced prototypes of a new generation of digital electron detectors by 2012. The innovative chips were sufficiently resistant to radiation damage to enable the direct detection of electrons, which improved the efficiency of detection to around half of the incoming electrons²⁰. Moreover, the modernized electronics that surround these direct electron detectors facilitated fast image capture, much like the burst mode of contemporary photographic cameras. Counting individual electron events on the fastest camera available led to an even better efficiency of electron detection²¹. Fast image capture also addressed a problem that is associated with frozen hydrated samples: energy released by the incoming electrons causes movement inside the ice layer, which blurs the resulting images. Movies recorded during exposure of the sample to electrons could be processed to effectively remove the blurring effects^{22,23}, which facilitated structure determination to an unprecedented resolution from far fewer data than before^{21,24}.

Meanwhile, another main impediment to high-resolution cryo-EM-based structure determination had been solved. Many macromolecular machines adopt a range of conformations in solution, and the purification or formation of these protein complexes is seldom perfect. This means that samples prepared for cryo-EM often contain a variety of structures. When such mixtures are subjected to single-particle analysis, the 2D projections of a number of 3D structures therefore need to be separated. A general solution to the mixture problem was provided first by 3D maximum-likelihood classification algorithms^{25,26}. Because they incorporate a statistical description of the data, these methods are more robust to noise than those that were already in common use²⁷. Alternative approaches to classification soon followed, which resulted in methods that turned the mixture problem into an opportunity. Whereas the presence of numerous structures previously blurred cryo-EM structures, valuable insight into protein dynamics could now be obtained from a single experiment²⁸.

When digital electron detectors became available commercially in 2013, the cryo-EM field was poised for a revolution. The unprecedented image quality that arose from these detectors enabled orientations to be assigned with greater accuracy. The ability to separate particles from distinct structural states was also improved. Structures that heralded this era include the coenzyme F₄₂₀-reducing hydrogenase²⁹, the mammalian ion channel transient receptor potential cation channel subfamily V member 1 (TRPV1) (ref. 30) and the large subunit of the mitochondrial ribosome in yeast³¹. The extension of maximum-likelihood methods into an empirical Bayesian approach made image processing more accessible to non-experts because crucial parameters no longer needed to be tuned; they were estimated from the data instead³². Moreover, the automated data-acquisition procedures that were developed originally for CCDs facilitated the recording of large datasets, from which the best particles could then be selected using image classification. As a result, in the past 3 years many research groups have solved cryo-EM structures to atomic resolution for a wide range of samples. Complexes with molecular weights of less than 200 kDa have become feasible targets for structure determination^{33,34}, and the achievable limits of resolution now extend to below 3 Å (refs 2, 35–37) and might even surpass 2 Å in favourable cases³⁴.

Opportunities for structure determination

Many of the newly determined structures represent proteins that are naturally embedded in membranes (Fig. 2). Such proteins are difficult to purify in solution because their hydrophobic membrane-spanning domains must be stabilized with detergents. Detergents also make crystallization notoriously difficult to achieve.

Consequently, membrane proteins are a blind spot for structural biology (Fig. 1), which is unfortunate because approximately half of all known small-molecule drugs bind to these proteins. However, structure determination through cryo-EM does not require crystallization. Instead, it is possible to image membrane proteins directly that have been solubilized in detergents or amphipols³⁸, or stabilized in a lipid environment using nanodiscs formed within the scaffold of an amphipathic protein belt³⁹ or the saposin–lipoprotein system⁴⁰. The first membrane-protein structure to be solved using the new cryo-EM technology was TRPV1 (ref. 30). This protein is responsible for the burning sensation that chilli peppers impart and it is an important drug target for pain. Its structure was solved first in an empty state that had been solubilized in amphipols. A complex of TRPV1 with the spider toxin DkTx and a small molecule that is similar to capsaicin, the active component of chilli peppers, was also determined⁴¹. An improved TRPV1 structure that was determined in the more natural environment of a nanodisc enabled the visualization of lipid substrates and provided more detailed mechanistic insight into TRPV1 function⁴². After the initial TRPV1 structure

was determined, the structures of many other medically relevant ion channels were published, including the voltage-dependent calcium channel Ca_v1.1 (ref. 43), a sodium–potassium channel⁴⁴ and the glycine receptor⁴⁵. Together, these structures have yielded a wealth of information on how cells regulate ion transport across membranes.

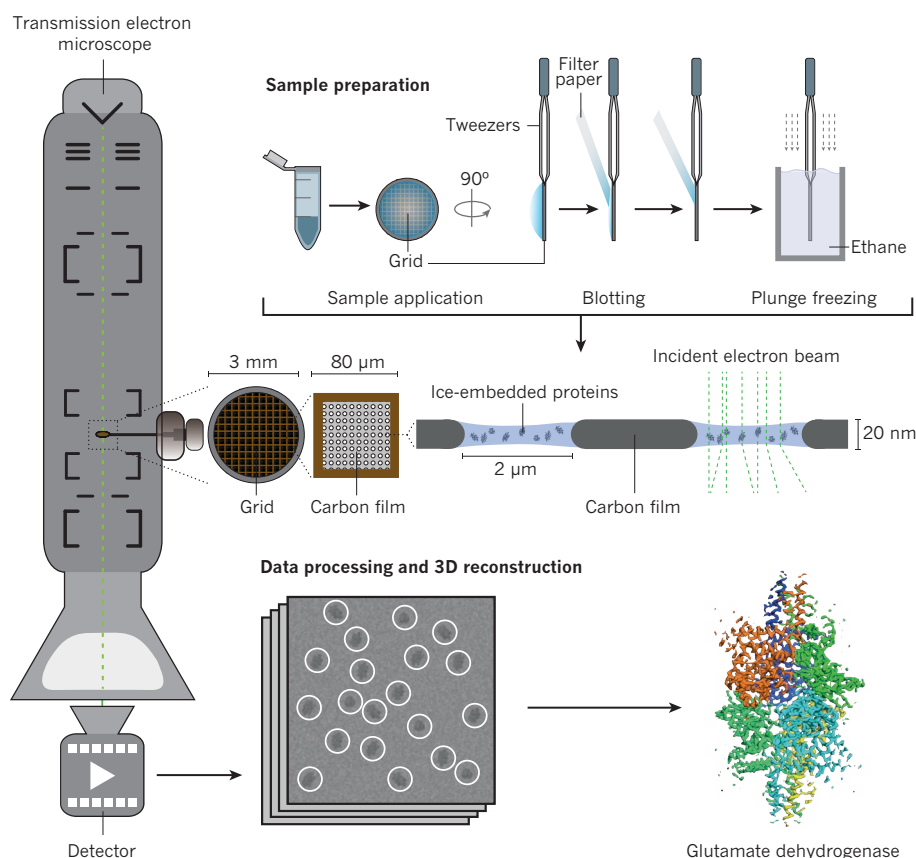
Another advantage of cryo-EM is that flexible regions of proteins do not impede structure determination. To facilitate crystallization, flexible loops or sugars are typically removed using complicated protein-engineering approaches⁴⁶. However, fully glycosylated wild-type proteins can be used for cryo-EM. For example, the structure of the human γ -secretase complex was solved despite it containing at least 11 sugar chains and a long disordered loop³³. This membrane-embedded protease generates amyloid- β peptides that aggregate in the brains of people with Alzheimer's disease. With an ordered mass of about 130 kDa, human γ -secretase is the smallest cryo-EM structure to be determined at a resolution below 3.5 Å. Another example is the heavily glycosylated structure of the human immunodeficiency virus type 1 (HIV-1) envelope glycoprotein trimer. This protein recognizes receptors on the surface of immune cells and mediates the

BOX 1

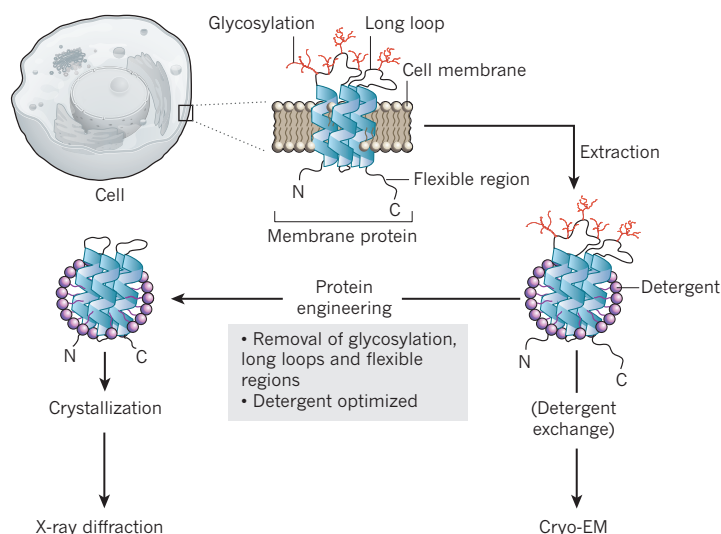
Cryo-EM structure determination

Protein structure determination through cryo-EM involves several stages: sample grid preparation, data collection and data processing followed by 3D reconstruction (Box Fig.). A few microlitres of a purified protein solution are applied to a grid composed of a perforated carbon film and the excess liquid is blotted away. The grid is plunged into liquid ethane, which flash-freezes the sample and embeds the particles in vitreous ice. The grid is stored in

liquid nitrogen until it is transferred to the transmission electron microscope. Two-dimensional images of proteins in various orientations within the grid's holes are then captured by the microscope's detector. The data are processed to combine images of the same protein into a 3D reconstruction of the protein's structure. The enzyme glutamate dehydrogenase³⁴ is presented as an example of a structure determined through cryo-EM.



a Extraction of membrane proteins



b Examples of membrane proteins

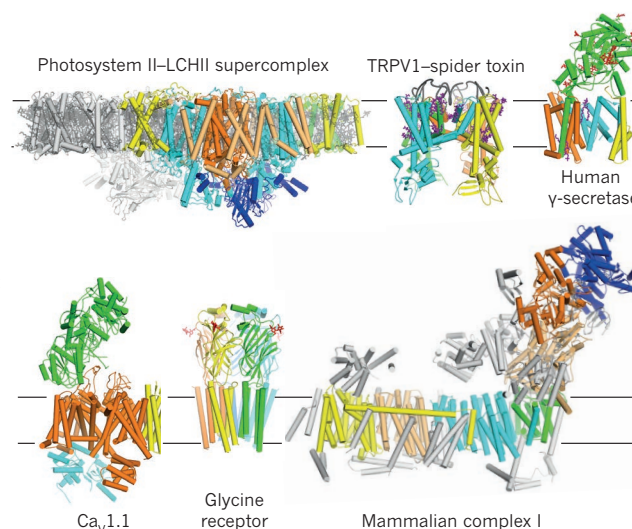


Figure 2 | Membrane protein structural biology. **a**, Proteins that are embedded in biological membranes must be extracted from the membrane using detergent solubilization techniques before their structures can be determined. The crystallization of membrane proteins for X-ray diffraction-based structure determination is difficult and often requires protein engineering to remove long loops, flexible regions and glycosylation. However, cryo-EM structure determination enables membrane proteins to be imaged directly in

entry of HIV-1. Knowledge of the structure of the fully glycosylated native protein has provided important information that might help in the development of HIV-1 vaccines^{47,48}. Similarly, the structures of Ebola virus glycoprotein GP1 bound to neutralizing antibodies⁴⁹ and to the transporter protein Niemann–Pick C1 (NPC1) (ref. 50) provide clues about immunity to Ebola in humans and how the virus fuses its membrane with that of the host.

Complexes that have been purified from native sources may also be suitable for cryo-EM-based structure determination. This is because wild-type proteins can be used and only micrograms of purified protein are required. Consequently, the structures of various large membrane-protein complexes that had resisted crystallographic studies have been determined, including the ryanodine receptor^{51–53}, the inositol triphosphate receptor⁵⁴, several glutamate receptors^{55–58}, mammalian complex I (ref. 59), photosystem II–light harvesting supercomplex (PSII–LHCII) (ref. 60) and ATP synthases^{61,62}.

The possibility of determining the structures of large complexes from native sources has also had a huge impact on the characterization of soluble macromolecular machines (Fig. 3). In the past 2 years, more than 50 high-resolution ribosome structures have been determined, which has yielded a wealth of information on the control of protein biosynthesis. For example, studies of both mammalian^{63,64} and yeast³¹ mitochondrial ribosomes have revealed how coevolution with the mitochondrial genome has affected their structures. Other macromolecular machines have also now become amenable to structural studies, as exemplified by the structures that have emerged of the inflammasome^{65,66}, the spliceosome^{67–69}, the signalosome^{70,71}, the exosome⁷², the anaphase-promoting complex^{73,73}, the 26S proteasome^{75,76}, the SNARE (soluble NSF attachment protein receptor) complex^{77,78}, the dynein–dynactin complex^{79,80}, the chaperone heat shock protein 90 (ref. 81) and the serine/threonine protein kinase mTOR^{82,83}. These all change conformation, rearrange their subunits and bind various partners and substrates during their assembly and working cycles. Cryo-EM enables the study of these often short-lived conformations and interactions, which are difficult to isolate or stabilize biochemically.

detergents or more natural environments such as nano-discs. The structures of glycosylated membrane proteins and very large membrane complexes can be determined through cryo-EM. **b**, Examples of membrane protein structures that were determined through cryo-EM: the photosystem II–light harvesting II supercomplex⁶⁰, TRPV1 in a complex with the spider toxin DsTx⁴², human γ -secretase³³, the voltage-gated calcium channel $\text{Ca}_v1.1$ (ref. 43), the glycine receptor⁴⁵ and mammalian complex I (ref. 59) are shown as examples.

Nuclear complexes that act on DNA and RNA form another group of molecular machines that is difficult to analyse conventionally. Both the proteins and the oligonucleotide substrates in these complexes tend to be highly dynamic molecules that engage in transient interactions with each other. Consequently, it has proven challenging to make or purify many large nuclear complexes. In a showcase of the potential of cryo-EM for studying highly dynamic complexes, various cryo-EM-determined structures are starting to shed light on the molecular details of some of the most fundamental processes of life. For example, DNA replication machinery structures have unveiled some of the molecular details of how genomes are copied^{84–87}, and a number of structures of RNA polymerases have led to insight into how DNA is transcribed into RNA^{88–92}. RNA polymerases have even been studied *in situ* by cryo-EM, inside the capsid of double-stranded RNA viruses^{93,94}. Other examples of cryo-EM studies of nuclear complexes include: a structure that provides a fresh understanding of genetic recombination⁹⁵; structures that show how the retroviral recombination machinery engages with the host nucleosome to insert its DNA^{96,97}; a structure of the bacterial group II intron that reveals how this mobile DNA element catalyses self-splicing in conjunction with a small intron-encoded protein⁹⁸; and structures that have aided our understanding of the mechanisms of type II and type III clustered regularly interspaced short palindromic repeat (CRISPR) systems^{99,100}.

In almost all of these studies, image classification played a crucial part in selecting structurally homogeneous subsets of particles for structure determination. For many structures, only a small fraction of the particles in the initial dataset is selected for use in the calculation of the final map. Most particles in cryo-EM datasets are unsuitable in some way for high-resolution structure determination and image classification enables only the best to be selected. Moreover, much like man-made machines, macromolecular machines make use of movements of parts relative to each other in their function. Whereas dynamic complexes would need to be trapped in a single state to facilitate crystallization, the new image classification algorithms offer the unique opportunity to visualize the full conformational freedom of such complexes in a single experiment (Fig. 4). A

striking example of this is the membrane-embedded ATP synthase, which acts as a molecular turbine, converting a proton flux into rotation to synthesize molecules of ATP, the energy currency of the cell. Image classification revealed the presence of three rotated states of the machine from a single cryo-EM dataset⁶¹. In another example, the imaging of actively translating human polysomes, which consist of multiple ribosomes bound to a single molecule of mRNA, followed by extensive image classification led to the production of structural snapshots of the entire translation cycle¹⁰¹.

The future of cryo-EM

Several challenges must be overcome for cryo-EM-based structure determination to continue its transformative growth in structural biology. An immediate concern is the elevated cost of cryo-EM. High-resolution structure determination is best performed using 300 kV electron microscopes, which cost in excess of US\$5 million and are accompanied by expensive maintenance contracts. Microscopes that operate at 200 kV are cheaper and can also be used to produce atomic-resolution structures^{102,103}; but their cost is still on the order of millions of dollars. To facilitate broad access to high-end microscopes, regional or national cryo-EM facilities are quickly gaining in popularity¹⁰⁴. However, the optimization of samples requires access to an electron microscope on a daily or weekly basis, which is not practical when using centralized facilities. In principle, expensive high-voltage machines are not needed for sample screening. But the cheaper microscopes that are available at present, which often operate at 100–120 kV, are unable to generate electrons that are all in-phase; they lack the coherence that is required to visualize particles with molecular weights below several hundreds of thousands of daltons. Consequently, there is an urgent need to develop affordable screening microscopes with more coherent electron sources.

The cost of storing and processing large volumes of data is also a problem. Automated data acquisition on a high-end microscope can yield several terabytes of images every day, and processing times can reach hundreds of thousands of computing core hours per dataset. To avoid the need to buy and maintain costly high-performance computing infrastructure, alternative solutions such as cloud computing¹⁰⁵ and the implementation of image-processing algorithms on cheaper graphics processing units (GPUs) are being explored^{106,107}.

As well as reducing costs and increasing the accessibility of high-end cryo-EM instruments, there is ample scope for improving the performance of microscope hardware. For example, progress has been achieved by increasing the efficiency of electron detection from about 30% when film is used to 50% for direct electron detectors. Such detectors could even be improved further. An important aspect of efficient electron detection is the ability to count individual electrons as they hit the detector, which requires fast read-outs^{21,108}. Of the three commercially available detectors, at present only one is fast enough to enable counting; another needs to lower the intensity of the electron beam to avoid flooding the chip with too many events. The use of more modern technology, such as faster electronics and improved chip design, could enable the production of detectors with efficiencies of up to 90%.

Noise in images can be decreased further through the use of energy filters. These optical devices remove electrons that have lost part of their energy in the sample and can no longer contribute constructively to the image. Energy filters were originally considered to be most useful for thick samples such as whole cells, in which more electrons lose energy while passing through the sample. However, some of the highest resolution and smallest structures that are available were reconstructed from energy-filtered images, which indicates that the removal of these electrons is also beneficial for thinner samples^{2,33,34,36}.

Optical devices known as phase plates are another promising development for microscope hardware. They produce a difference in the phase of the scattered and the unscattered electron waves. This

generates contrast that is improved by up to an order of magnitude in images at low resolution — of particular interest when imaging complexes that are too small to yield enough contrast with existing optics¹⁰⁹. Aberration correctors, a further type of optical device, can yield better images, in particular at higher resolutions. These devices counteract the effects of imperfections in the optical system of the microscope through complicated combinations of lenses. Although not needed at present to reach resolutions of 3 Å or lower, such correctors might be helpful when higher resolutions of around 2 Å must be achieved¹¹⁰.

The refinement of sample preparation methods provides another important opportunity for enhancing cryo-EM-based structure determination. Although movie processing has progressed enough to correct for beam-induced motion in samples, large movements of particles during the early stages of electron exposure are often too fast to be corrected^{21,29,111}. This places a considerable constraint on the determination of structures at high resolution because most of the high-resolution information is destroyed by radiation damage during this early period¹¹². Developments in sample preparation for cryo-EM aim to reduce or stop this motion, for example by replacing both the copper grid and the amorphous holey carbon film with gold¹¹³, or by using films made of graphene or graphene oxide^{114,115}.

Requirements that concern the quantity and purity of samples can also be modified to boost cryo-EM outcomes. At present, several microlitres of a purified sample of protein, typically at a concentration of 0.1–5.0 μmol per litre, are needed to prepare a single cryo-EM grid. Purification of the sample on the grid itself, through the use of specifically adhered affinity tags¹¹⁶, can relax demands on sample concentration and purity. And because less than 0.1% of the sample volume will remain on the grid after blotting, and only a fraction of the grid's surface is typically used for data acquisition, savings in sample volumes of multiple orders of magnitude are possible. Investigations that have pursued this direction include spraying picolitre-sized droplets of samples on cryo-EM grids¹¹⁷. Similar methods can

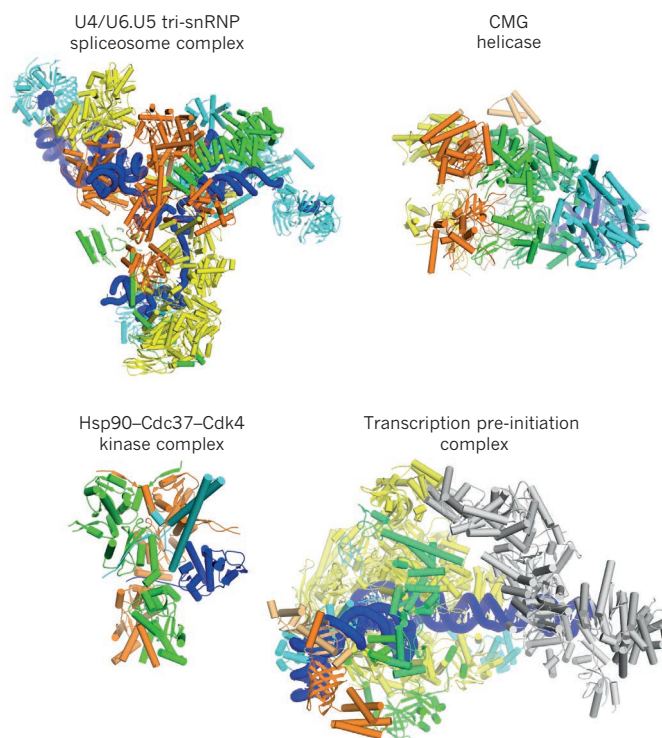


Figure 3 | Soluble macromolecular machines. Examples of cryo-EM structures are shown for the U4/U6.U5 tri-snRNP spliceosomal complex in yeast⁶⁷; the eukaryotic replicative CMG helicase⁸⁶; the human Hsp90–Cdc37–Cdk4 kinase complex⁸¹; and the human transcription pre-initiation complex²¹.

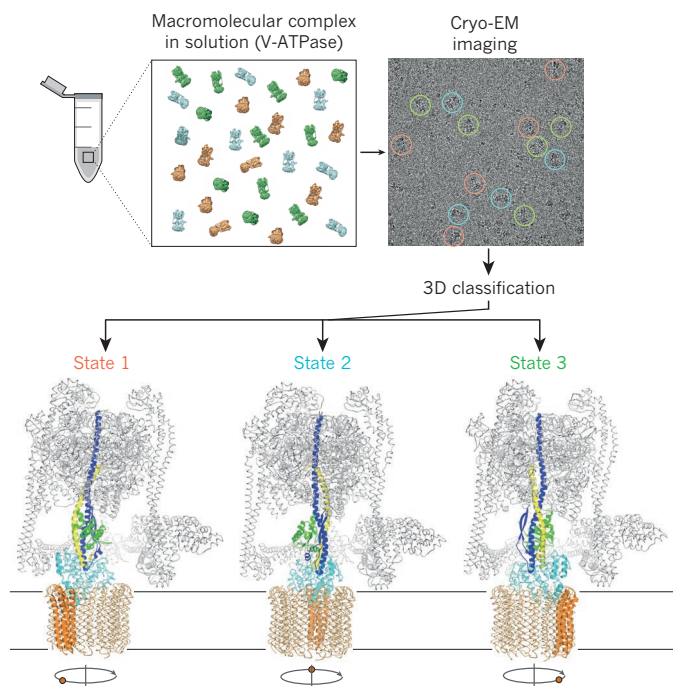


Figure 4 | Image classification enables the study of macromolecular dynamics. A mixture containing a macromolecular complex in distinct conformations or compositions (shown as red, blue and green objects) can be imaged directly using cryo-EM techniques. Three-dimensional image classification is then used to obtain structures for each state of the complex. Such classification therefore enables the functional cycles of dynamic molecular machines such as the eukaryotic V-ATPase⁶¹ to be characterized from a single experiment.

be used in time-resolved studies in which multiple protein components are mixed together in a precisely timed manner¹¹⁸. When combined with modern microfluidic systems and the ability to tackle mixtures through image classification, this could enable reactions to be followed on a millisecond timescale¹¹⁹.

In software development, superior image-formation models might be necessary to calculate structures of higher resolution. Improved image-processing algorithms are also needed to deal with complicated, multicomponent mixtures and to identify less common states. In particular, the presence of numerous continuously flexing domains still represents a challenge. Alternative procedures that describe the ensemble of structures in a dataset (instead of dividing the data into subsets) have the potential to reveal the conformational complexity of highly dynamic molecular machines^{67,120}.

Prospects

Each of the developments described in this Review has the potential to enlarge the scope of cryo-EM-based single-particle analysis and will lead to the collection of images with higher signal-to-noise ratios. Together with improved image-processing algorithms, such images will yield higher resolution structures of smaller complexes and of more complicated mixtures than is possible at present.

The ability to visualize proteins complexes with molecular weights of around 100 kDa will widen the applicability of the technique to many more drug targets. For example, G-protein-coupled receptors are sought-after targets for the pharmaceutical industry. In complex with conformation-specific antibodies or heterotrimeric G proteins, these receptors should become amenable to structure determination by cryo-EM. With the potential to achieve resolutions well beyond 3 Å (ref. 34) for complexes that are bound to drug candidates¹²¹, cryo-EM structure-based drug design will become routine for targets that, at present, are extremely difficult to characterize using alternative techniques.

The ability to obtain structures from small subsets of the data through image processing, possibly in combination with sample preparation on the nanolitre scale, will also expand the applicability of cryo-EM to complexes with low abundance in the cell. Using time-resolved methods, even transient complexes and intermediate conformational states can be studied, providing unprecedented insight into the function of large macromolecular machines. For example, the structural characterization at the molecular (or even the atomic) level of extremely large and flexible machines such as the nuclear pore complex¹²², as well as the characterization of complexes involved in the organization of chromatin¹²³, is now on the horizon.

Yet macromolecules do not act in isolation. The cell is densely packed with many molecules that interact to form highly intricate networks that are finely tuned to keep it alive. The detailed structural information about individual complexes that can be obtained through single-particle analysis needs to be understood in the context of the complexes' cellular environment. Excitingly, imaging with electrons could also provide unique opportunities in this task. For example, cryo-electron tomography can be used to build 3D reconstructions of frozen cells by taking multiple images of the same sample at different tilt angles inside the electron microscope¹²⁴. Radiation damage is still a major limitation, but the resolution of the substructures that are repeated in the cell can be increased by applying averaging methods¹²⁵. The tomographic approaches will benefit from the same developments in hardware and software that are driving forwards cryo-EM-based single-particle analysis. These techniques could bridge the gap between biophysics and cell biology and provide a road map for understanding of the cell at a molecular level. ■

Received 11 April; accepted 15 July 2016.

1. Alberts, B. The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **92**, 291–294 (1998).
2. Banerjee, S. *et al.* 2.3 Å resolution cryo-EM structure of human p97 and mechanism of allosteric inhibition. *Science* **351**, 871–875 (2016).
3. Kühlbrandt, W. The resolution revolution. *Science* **343**, 1443–1444 (2014).
4. Taylor, K. A. & Glaeser, R. M. Electron diffraction of frozen, hydrated protein crystals. *Science* **186**, 1036–1037 (1974).
5. Dubochet, J., Chang, J.-J., Freeman, R., Lepault, J. & McDowell, A. W. Frozen aqueous suspensions. *Ultramicroscopy* **10**, 55–62 (1982).
6. Frank, J., Verschoor, A. & Boublik, M. Computer averaging of electron micrographs of 40S ribosomal subunits. *Science* **214**, 1353–1355 (1981).
7. Jiang, W. *et al.* Backbone structure of the infectious ε15 virus capsid revealed by electron cryomicroscopy. *Nature* **451**, 1130–1134 (2008).
8. Yu, X., Jin, L. & Zhou, Z. H. 3.88 Å structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy. *Nature* **453**, 415–419 (2008).
9. Zhang, X. *et al.* Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. *Proc. Natl Acad. Sci. USA* **105**, 1867–1872 (2008).
10. Villa, E. *et al.* Ribosome-induced changes in elongation factor Tu conformation control GTP hydrolysis. *Proc. Natl Acad. Sci. USA* **106**, 1063–1068 (2009).
11. Schuette, J.-C. *et al.* GTPase activation of elongation factor EF-Tu by the ribosome during decoding. *EMBO J.* **28**, 755–765 (2009).
12. Seidelt, B. *et al.* Structural insight into nascent polypeptide chain-mediated translational stalling. *Science* **326**, 1412–1415 (2009).
13. Agard, D., Cheng, Y., Glaeser, R. M. & Subramaniam, S. in *Advances in Imaging and Electron Physics* Vol. 185 (ed. Hawkes, P. W.) Ch. 2, 113–137 (Elsevier, 2014).
14. Frank, J. Generalized single-particle cryo-EM—a historical perspective. *Microscopy* **65**, 3–8 (2016).
15. Vinokumar, K. R. & Henderson, R. Single particle electron cryomicroscopy: trends, issues and future perspective. *Q. Rev. Biophys.* **49**, e13 (2016).
16. Henderson, R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.* **28**, 171–193 (1995).
17. Krivanek, O. L. & Mooney, P. E. Applications of slow-scan CCD cameras in transmission electron microscopy. *Ultramicroscopy* **49**, 95–108 (1993).
18. Potter, C. S. *et al.* Legion: a system for fully automated acquisition of 1000 electron micrographs a day. *Ultramicroscopy* **77**, 153–161 (1999).
19. McMullan, G., Chen, S., Henderson, R. & Faruqi, A. R. Detective quantum efficiency of electron area detectors in electron microscopy. *Ultramicroscopy* **109**, 1126–1143 (2009).
20. McMullan, G., Faruqi, A. R., Clare, D. & Henderson, R. Comparison of optimal performance at 300 keV of three direct electron detectors for use in low dose electron microscopy. *Ultramicroscopy* **147**, 156–163 (2014).

A comparison of the three commercially available direct electron detectors; all were shown to perform better than photographic film.

21. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* **10**, 584–590 (2013).
22. Brilot, A. F. *et al.* Beam-induced motion of vitrified specimen on holey carbon film. *J. Struct. Biol.* **177**, 630–637 (2012).
23. Campbell, M. G. *et al.* Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. *Structure* **20**, 1823–1828 (2012).
24. Bai, X.-C., Fernandez, I. S., McMullan, G. & Scheres, S. H. W. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *eLife* **2**, e00461 (2013).
25. Scheres, S. H. W. *et al.* Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nature Methods* **4**, 27–29 (2007).
26. Lyumkis, D., Brilot, A. F., Theobald, D. L. & Grigorieff, N. Likelihood-based classification of cryo-EM images using FREALIGN. *J. Struct. Biol.* **183**, 377–388 (2013).
27. Sigworth, F. J. A. Maximum-likelihood approach to single-particle image refinement. *J. Struct. Biol.* **122**, 328–339 (1998).
28. Clare, D. K. *et al.* ATP-triggered conformational changes delineate substrate-binding and -folding mechanics of the GroEL chaperonin. *Cell* **149**, 113–123 (2012).
29. Allegretti, M., Mills, D. J., McMullan, G., Kühlbrandt, W. & Vonck, J. Atomic model of the F_{420} -reducing [NiFe] hydrogenase by electron cryo-microscopy using a direct electron detector. *eLife* **3**, e01963 (2014).
- Refs 29–31 each present one of the first three high-resolution structures that heralded the resolution revolution in cryo-EM-based structure determination.**
30. Liao, M., Cao, E., Julius, D. & Cheng, Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* **504**, 107–112 (2013).
31. Amunts, A. *et al.* Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343**, 1485–1489 (2014).
32. Scheres, S. H. W. A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415**, 406–418 (2012).
- A statistical framework for the classification and high-resolution refinement of cryo-EM structures that reduces the need for expert supervision.**
33. Bai, X.-C. *et al.* An atomic structure of human γ -secretase. *Nature* **525**, 212–217 (2015).
34. Merk, A. *et al.* Breaking cryo-EM resolution barriers to facilitate drug discovery. *Cell* **165**, 1698–1707 (2016).
- The highest reported resolution (1.8 Å) of a single-particle reconstruction, at present.**
35. Campbell, M. G., Veesler, D., Cheng, A., Potter, C. S. & Carragher, B. 2.8 Å resolution reconstruction of the *Thermoplasma acidophilum* 20S proteasome using cryo-electron microscopy. *eLife* **4**, e06380 (2015).
36. Bartesaghi, A. *et al.* 2.2 Å resolution cryo-EM structure of β -galactosidase in complex with a cell-permeant inhibitor. *Science* **348**, 1147–1151 (2015).
37. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**, e06980 (2015).
38. Althoff, T., Mills, D. J., Popot, J. L. & Kühlbrandt, W. Arrangement of electron transport chain components in bovine mitochondrial supercomplex I₁III₂IV₁. *EMBO J.* **30**, 4652–4664 (2011).
39. Bayburt, T. H. & Sligar, S. G. Membrane protein assembly into nanodiscs. *FEBS Lett.* **584**, 1721–1727 (2010).
40. Fraumenfeld, J. *et al.* A saposin-lipoprotein nanoparticle system for membrane proteins. *Nature Methods* **13**, 345–351 (2016).
41. Cao, E., Liao, M., Cheng, Y. & Julius, D. TRPV1 structures in distinct conformations reveal activation mechanisms. *Nature* **504**, 113–118 (2013).
42. Gao, Y., Cao, E., Julius, D. & Cheng, Y. TRPV1 structures in nanodiscs reveal mechanisms of ligand and lipid action. *Nature* **534**, 347–351 (2016).
43. Wu, J. *et al.* Structure of the voltage-gated calcium channel $Ca_v1.1$ complex. *Science* **350**, aad2395 (2015).
44. Hite, R. K. *et al.* Cryo-electron microscopy structure of the Slo2.2 Na^+ -activated K^+ channel. *Nature* **527**, 198–203 (2015).
45. Du, J., Lü, W., Wu, S., Cheng, Y. & Gouaux, E. Glycine receptor mechanism elucidated by electron cryo-microscopy. *Nature* **526**, 224–229 (2015).
46. Baker, M. Making membrane proteins for structures: a trillion tiny tweaks. *Nature Methods* **7**, 429–434 (2010).
47. Lyumkis, D. *et al.* Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. *Science* **342**, 1484–1490 (2013).
48. Lee, J. H., Ozorowski, G. & Ward, A. B. Cryo-EM structure of a native, fully glycosylated, cleaved HIV-1 envelope trimer. *Science* **351**, 1043–1048 (2016).
49. Misasi, J. *et al.* Structural and molecular basis for Ebola virus neutralization by protective human antibodies. *Science* **351**, 1343–1346 (2016).
50. Gong, X. *et al.* Structural insights into the Niemann–Pick C1 (NPC1)-mediated cholesterol transfer and Ebola infection. *Cell* **165**, 1467–1478 (2016).
51. Yan, Z. *et al.* Structure of the rabbit ryanodine receptor RyR1 at near-atomic resolution. *Nature* **517**, 50–55 (2015).
52. Efremov, R. G., Leitner, A., Aebersold, R. & Raunser, S. Architecture and conformational switch mechanism of the ryanodine receptor. *Nature* **517**, 39–43 (2015).
53. Zalk, R. *et al.* Structure of a mammalian ryanodine receptor. *Nature* **517**, 44–49 (2015).
54. Fan, G. *et al.* Gating machinery of $InsP_3R$ channels revealed by electron cryomicroscopy. *Nature* **527**, 336–341 (2015).
55. Tajima, N. *et al.* Activation of NMDA receptors and the mechanism of inhibition by ifenprodil. *Nature* **534**, 63–68 (2016).
56. Zhu, S. *et al.* Mechanism of NMDA receptor inhibition and activation. *Cell* **165**, 704–714 (2016).
57. Herguedas, B. *et al.* Structure and organization of heteromeric AMPA-type glutamate receptors. *Science* **352**, aad3873 (2016).
58. Meyerson, J. R. *et al.* Structural mechanism of glutamate receptor activation and desensitization. *Nature* **514**, 328–334 (2014).
59. Vinothkumar, K. R., Zhu, J. & Hirst, J. Architecture of mammalian respiratory complex I. *Nature* **515**, 80–84 (2014).
60. Wei, X. *et al.* Structure of spinach photosystem II–LHCII supercomplex at 3.2 Å resolution. *Nature* **534**, 69–74 (2016).
61. Zhao, J., Benlekhir, S. & Rubinstein, J. L. Electron cryomicroscopy observation of rotational states in a eukaryotic V-ATPase. *Nature* **521**, 241–245 (2015).
- An example of how image classification can reveal numerous functional states of dynamic molecular machines from a single experiment.**
62. Allegretti, M. *et al.* Horizontal membrane-intrinsic α -helices in the stator α -subunit of an F-type ATP synthase. *Nature* **521**, 237–240 (2015).
63. Amunts, A., Brown, A., Toots, J., Scheres, S. H. W. & Ramakrishnan, V. The structure of the human mitochondrial ribosome. *Science* **348**, 95–98 (2015).
64. Greber, B. J. *et al.* The complete structure of the 55S mammalian mitochondrial ribosome. *Science* **348**, 303–308 (2015).
65. Zhang, L. *et al.* Cryo-EM structure of the activated NAIIP2-NLRC4 inflammasome reveals nucleated polymerization. *Science* **350**, 404–409 (2015).
66. Hu, Z. *et al.* Structural and biochemical basis for induced self-propagation of NLRC4. *Science* **350**, 399–404 (2015).
67. Nguyen, T. H. D. *et al.* Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. *Nature* **530**, 298–302 (2016).
68. Agafonov, D. E. *et al.* Molecular architecture of the human U4/U6.U5 tri-snRNP. *Science* **351**, 1416–1420 (2016).
69. Wan, R. *et al.* The 3.8 Å structure of the U4/U6.U5 tri-snRNP: insights into spliceosome assembly and catalysis. *Science* **351**, 466–475 (2016).
70. Mosadeghi, R. *et al.* Structural and kinetic analysis of the COP9–signalosome activation and the cullin–RING ubiquitin ligase deneddylation cycle. *eLife* **5**, e12102 (2016).
71. Cavadini, S. *et al.* Cullin–RING ubiquitin E3 ligase regulation by the COP9 signalosome. *Nature* **531**, 598–603 (2016).
72. Liu, J.-J. *et al.* CryoEM structure of yeast cytoplasmic exosome complex. *Cell Res.* **26**, 822–837 (2016).
73. Chang, L., Zhang, Z., Yang, J., McLaughlin, S. H. & Barford, D. Atomic structure of the APC/C and its mechanism of protein ubiquitination. *Nature* **522**, 450–454 (2015).
74. Zhang, S. *et al.* Molecular mechanism of APC/C activation by mitotic phosphorylation. *Nature* **533**, 260–264 (2016).
75. Schweitzer, A. *et al.* Structure of the human 26S proteasome at a resolution of 3.9 Å. *Proc. Natl Acad. Sci. USA* **113**, 7816–7821 (2016).
76. Dambacher, C. M., Worden, E. J., Herzik, M. A., Martin, A. & Lander, G. C. Atomic structure of the 26S proteasome lid reveals the mechanism of deubiquitinase inhibition. *eLife* **5**, e13027 (2016).
77. Zhou, Q. *et al.* Cryo-EM structure of SNAP-SNARE assembly in 20S particle. *Cell Res.* **25**, 551–560 (2015).
78. Zhao, M. *et al.* Mechanistic insights into the recycling machine of the SNARE complex. *Nature* **518**, 61–67 (2015).
79. Chowdhury, S., Ketcham, S. A., Schroer, T. A. & Lander, G. C. Structural organization of the dynein–dynactin complex bound to microtubules. *Nature Struct. Mol. Biol.* **22**, 345–347 (2015).
80. Urnauvicius, L. *et al.* The structure of the dynactin complex and its interaction with dynein. *Science* **347**, 1441–1446 (2015).
81. Verba, K. A. *et al.* Atomic structure of Hsp90–Cdc37–Cdk4 reveals that Hsp90 traps and stabilizes an unfolded kinase. *Science* **352**, 1542–1547 (2016).
82. Aylett, C. H. S. *et al.* Architecture of human mTOR complex 1. *Science* **351**, 48–52 (2016).
83. Baretić, D., Berndt, A., Ohashi, Y., Johnson, C. M. & Williams, R. L. Tor forms a dimer through an N-terminal helical solenoid with a complex topology. *Nature Commun.* **7**, 11016 (2016).
84. Fernández-Leiro, R., Conrad, J., Scheres, S. H. & Lamers, M. H. Cryo-EM structures of the *E. coli* replicative DNA polymerase reveal its dynamic interactions with the DNA sliding clamp, exonuclease and τ . *eLife* **4**, e11134 (2015).
85. Li, N. *et al.* Structure of the eukaryotic MCM complex at 3.8 Å. *Nature* **524**, 186–191 (2015).
86. Yuan, Z. *et al.* Structure of the eukaryotic replicative CMG helicase suggests a pumpjack motion for translocation. *Nature Struct. Mol. Biol.* **23**, 217–224 (2016).
87. Abid Ali, F. *et al.* Cryo-EM structures of the eukaryotic replicative helicase bound to a translocation substrate. *Nature Commun.* **7**, 10708 (2016).
88. Hoffmann, N. A. *et al.* Molecular structures of unbound and transcribing RNA polymerase III. *Nature* **528**, 231–236 (2015).
89. Murakami, K. *et al.* Structure of an RNA polymerase II preinitiation complex. *Proc. Natl Acad. Sci. USA* **112**, 13543–13548 (2015).
90. Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J. M. & Cramer, P. Structure of transcribing mammalian RNA polymerase II. *Nature* **529**, 551–554 (2016).
91. He, Y. *et al.* Near-atomic resolution visualization of human transcription promoter opening. *Nature* **533**, 359–365 (2016).
92. Plaschka, C. *et al.* Transcription initiation complex structures elucidate DNA opening. *Nature* **533**, 353–358 (2016).
93. Zhang, X. *et al.* *In situ* structures of the segmented genome and RNA polymerase complex inside a dsRNA virus. *Nature* **527**, 531–534 (2015).
94. Liu, H. & Cheng, L. Cryo-EM shows the polymerase structures and a nonspooled genome within a dsRNA virus. *Science* **349**, 1347–1350 (2015).

95. Ru, H. *et al.* Molecular mechanism of V(D)J recombination from synaptic RAG1–RAG2 complex structures. *Cell* **163**, 1138–1152 (2015); erratum **163**, 1807 (2015).
96. Maskell, D. P. *et al.* Structural basis for retroviral integration into nucleosomes. *Nature* **523**, 366–369 (2015).
97. Ballandras-Colas, A. *et al.* Cryo-EM reveals a novel octameric integrase structure for betaretroviral intasome function. *Nature* **530**, 358–361 (2016).
98. Qu, G. *et al.* Structure of a group II intron in complex with its reverse transcriptase. *Nature Struct. Mol. Biol.* **23**, 549–557 (2016).
99. Jiang, F. *et al.* Structures of a CRISPR–Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867–871 (2016).
100. Taylor, D. W. *et al.* Structures of the CRISPR–Cmr complex reveal mode of RNA target positioning. *Science* **348**, 581–585 (2015).
101. Behrmann, E. *et al.* Structural snapshots of actively translating human ribosomes. *Cell* **161**, 845–857 (2015).
102. Campbell, M. G. *et al.* Near-atomic resolution reconstructions using a mid-range electron microscope operated at 200kV. *J. Struct. Biol.* **188**, 183–187 (2014).
103. Liang, B. *et al.* Structure of the L protein of vesicular stomatitis virus from electron cryomicroscopy. *Cell* **162**, 314–327 (2015).
104. Saibil, H. R., Grünwald, K. & Stuart, D. I. A national facility for biological cryo-electron microscopy. *Acta Crystallogr. D* **71**, 127–135 (2015).
105. Cianfrocco, M. A. & Leschziner, A. E. Low cost, high performance processing of single particle cryo-electron microscopy data in the cloud. *eLife* **4**, e06664 (2015).
106. Schmeisser, M. *et al.* Parallel, distributed and GPU computing technologies in single-particle electron microscopy. *Acta Crystallogr. D* **65**, 659–671 (2009).
107. Kimanius, D., Forsberg, B. O., Scheres, S. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. Preprint at <http://dx.doi.org/10.1101/059717> (2016).
108. McMullan, G., Clark, A. T., Turchetta, R. & Faruqi, A. R. Enhanced imaging in low dose electron microscopy using electron counting. *Ultramicroscopy* **109**, 1411–1416 (2009).
109. Danev, R. & Baumeister, W. Cryo-EM single particle analysis with the Volta phase plate. *eLife* **5**, e13046 (2016).
110. Fischer, N. *et al.* Structure of the *E. coli* ribosome–EF-Tu complex at <3 Å resolution by Cs-corrected cryo-EM. *Nature* **520**, 567–570 (2015).
111. Scheres, S. H. Beam-induced motion correction for sub-megadalton cryo-EM particles. *eLife* **3**, e03665 (2014).
112. Stark, H., Zemlin, F. & Boettcher, C. Electron radiation damage to protein crystals of bacteriorhodopsin at different temperatures. *Ultramicroscopy* **63**, 75–79 (1996).
113. Russo, C. J. & Passmore, L. A. Ultrastable gold substrates for electron cryomicroscopy. *Science* **346**, 1377–1380 (2014).
114. Russo, C. J. & Passmore, L. A. Controlling protein adsorption on graphene for cryo-EM using low-energy hydrogen plasmas. *Nature Methods* **11**, 649–652 (2014).
115. Pantelic, R. S., Meyer, J. C., Kaiser, U., Baumeister, W. & Plitzko, J. M. Graphene oxide: a substrate for optimizing preparations of frozen-hydrated samples. *J. Struct. Biol.* **170**, 152–156 (2010).
116. Kelly, D. F., Dukovski, D. & Walz, T. New applications for affinity grids in preparing EM specimens. *Microsc. Microanal.* **16** (suppl. S2), 840–841 (2010).
117. Jain, T., Sheehan, P., Crum, J., Carragher, B. & Potter, C. S. Spotiton: a prototype for an integrated inkjet dispense and vitrification system for cryo-TEM. *J. Struct. Biol.* **179**, 68–75 (2012).
118. Unwin, N. Acetylcholine receptor channel imaged in the open state. *Nature* **373**, 37–43 (1995).
119. Chen, B. *et al.* Structural dynamics of ribosome subunit association studied by mixing-spraying time-resolved cryogenic electron microscopy. *Structure* **23**, 1097–1105 (2015).
120. Dashti, A. *et al.* Trajectories of the ribosome as a Brownian nanomachine. *Proc. Natl Acad. Sci. USA* **111**, 17492–17497 (2014).
121. Li, H. *et al.* Structure- and function-based design of *Plasmodium*-selective proteasome inhibitors. *Nature* **530**, 233–236 (2016).
122. Kosinski, J. *et al.* Molecular architecture of the inner ring scaffold of the human nuclear pore complex. *Science* **352**, 363–365 (2016).
123. Song, F. *et al.* Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science* **344**, 376–380 (2014).
124. Baumeister, W. Electron tomography: towards visualizing the molecular organization of the cytoplasm. *Curr. Opin. Struct. Biol.* **12**, 679–684 (2002).
125. Mahamid, J. *et al.* Visualizing the molecular sociology at the HeLa cell nuclear periphery. *Science* **351**, 969–972 (2016).

This work presents an exciting vision of how cryo-electron tomography and sub-tomogram averaging can bridge the gap between the structural biology of isolated complexes and cell biology.

Acknowledgements We thank Y. Cheng, D. Barford and R. Henderson for comments on an early version of this manuscript. S.H.W.S is funded by the UK Medical Research Council (MC_UP_A025_1013).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this paper at go.nature.com/2bqncxm. Correspondence should be addressed to S.H.W.S. (scheres@mrc-lmb.cam.ac.uk).

Mass-spectrometric exploration of proteome structure and function

Ruedi Aebersold^{1,2} & Matthias Mann^{3,4}

Numerous biological processes are concurrently and coordinately active in every living cell. Each of them encompasses synthetic, catalytic and regulatory functions that are, almost always, carried out by proteins organized further into higher-order structures and networks. For decades, the structures and functions of selected proteins have been studied using biochemical and biophysical methods. However, the properties and behaviour of the proteome as an integrated system have largely remained elusive. Powerful mass-spectrometry-based technologies now provide unprecedented insights into the composition, structure, function and control of the proteome, shedding light on complex biological processes and phenotypes.

Collectively, proteins catalyse and control essentially all cellular processes. They form a highly structured entity known as the proteome, the constituent proteins of which carry out their functions at specific times and locations in the cell, in physical or functional association with other proteins or biomolecules. A proliferating *Schizosaccharomyces pombe* cell contains about 60 million protein molecules, which have abundances that range from a few copies to 1.1 million copies per expressed gene¹. Across the species, proteins constitute about 50% of the dry mass of a cell and reach a remarkable total concentration of 2–4 million proteins per cubic micrometre or 100–300 mg per ml (ref. 2). The extensive proteome network of the cell adapts dynamically to external or internal (that is, genetic) perturbations and thereby defines the cell's functional state and determines its phenotypes. Describing and understanding the complete and quantitative proteome as well as its structure, function and dynamics is a central and fundamental challenge of biology.

Two strategies that differ in principle have been used to study the proteome and the molecular mechanisms that it mediates. Conventionally, specific proteins are isolated and then analysed with respect to their structure and function through the established methods of biochemistry and biophysics. But it has also become possible to perform large-scale, systematic measurements of proteomes to generate biological insights from the computational analysis of proteomic datasets, either on their own or in combination with other 'omics' types of data. Both approaches have been transformed fundamentally by the development of powerful mass-spectrometry-based methods. Such techniques have the capability to identify conclusively and quantify accurately almost any protein that has been expressed. They can also systematically identify and localize modified amino acids in the polypeptide chain as well as determine the composition, stoichiometry and topology of the subunits of multiprotein complexes and even contribute to determining their structure.

The annotated genome identifies the entire proteome of an organism. However, the literature has focused on the small fraction of the proteome for which measurement assays are readily available³. This set of intensely studied proteins has remained surprisingly constant over the past few decades. Robust mass-spectrometry-based methods now enable most proteins to be measured reliably, which vastly extends the range of the classic, mechanism-focused analyses of specific components of the proteome. They also make possible the systematic analysis of the proteome to an extent that had been predicted previously^{4,5}.

Underlying reasons for the success of mass spectrometry in proteomics include its inherent specificity of identification, the generic nature of the proteomics workflow and its potential for extreme sensitivity that, in principle, extends to the single ion. In practice, it has been challenging to realize the full potential of the technique, and ingenious ways of implementing mass spectrometry as a universal detector of protein identity, abundance, precise chemical state and cellular context and localization are still being devised. At present, no single mass-spectrometry-based system or method can determine by itself these diverse dimensions for proteome data.

This Review highlights the achievements of mass-spectrometry-based proteomics and the challenges that remain. Efforts to catalogue systematically the proteomes of an array of species and to transform these catalogues into highly specific assays that can quantify any component are described. The analysis of post-translational modifications is discussed, especially with regard to completeness of measurement and how the research community might assign functions to the tens of thousands of modified sites that have been discovered in the past decade. The state of mass spectrometry is reviewed in the context of the study of functional modules, in which components of the proteome come together stably or temporarily in complexes to carry out a biochemical function. Last, mass-spectrometry-based techniques that are capable of quantifying thousands of proteins across collections of large numbers of samples with a high degree of reproducibility are described; these generate large datasets that can be mined by statistical machine-learning tools to determine the state of the proteome and its response to perturbations. Such datasets start to uncover systemic malfunctions at the cellular and organismal levels in diseases that have been difficult to reach through classic protein-based or nucleic-acid-based research.

The identification and quantification of the proteome

The ability to identify reliably any component of the proteome is a requirement both for mechanistic, hypothesis-driven investigations and for large-scale, omics-type studies. A comprehensive and reliable mass-spectrometry-based proteome map is also a prerequisite for the development of targeted mass spectrometry techniques, as well as for data-independent acquisition (DIA) strategies (Fig. 1 and Box 1); these rely on information from pre-existing high-quality spectral libraries. The importance of accurate quantification in proteomics is hard to overstate,

¹Institute of Molecular Systems Biology, Department of Biology, ETH Zürich, 8093 Zürich, Switzerland. ²Faculty of Science, University of Zürich, 8093 Zürich, Switzerland. ³Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany. ⁴Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark.

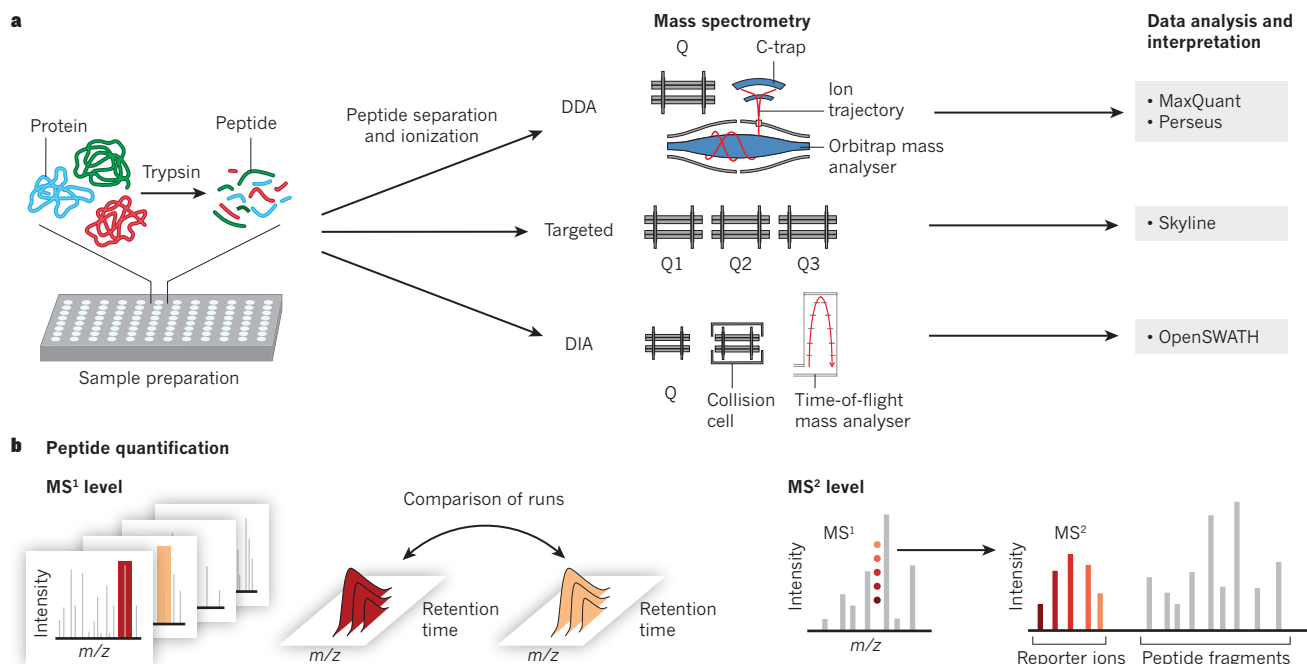


Figure 1 | Bottom-up proteomics workflows. **a**, All bottom-up proteomics workflows begin with a sample-preparation stage in which proteins are extracted and digested by a sequence-specific enzyme such as trypsin. Present methods of protein preparation are highly efficient and can be performed in 96-well plates with robotic assistance. Peptides are then separated by means of chromatography and electrosprayed, after which they are introduced into the vacuum of a mass spectrometer. Three classes of methods are shown. In DDA methods, a full spectrum of the peptides (at the MS¹ level) is acquired, followed by the collection of as many fragmentation spectra (at the MS² level) as possible, within a cycle time of about 1 second. A quadrupole–orbitrap mass analyser is depicted, although other types of analyser are also used in DDA. Results are interpreted using software packages such as MaxQuant¹⁰⁰ and the downstream Perseus environment¹⁰¹. In targeted analysis, a peptide of known mass-to-charge ratio (m/z) is selected in the first quadrupole, then the peptide is fragmented and several fragments are monitored over time. These transitions are multiplexed and their specificity is checked using software packages such as SkyLine¹⁰². In DIA methods, which are exemplified by sequential window acquisition of all theoretical fragment-ion spectra (SWATH)–MS¹⁰³, ranges of m/z values (that typically span 25 m/z units) are selected and peptides are fragmented, followed by the acquisition of the fragments in a time-of-flight mass spectrometer. The instrument rapidly and seamlessly cycles through the entire mass range within a few seconds. The multiplexed fragment spectra

are interpreted — often with the help of known fragment spectra from large spectral libraries — by software such as OpenSWATH¹⁰⁴. **b**, Peptide quantities can be determined at the MS¹ level by integrating the signal from peaks of the precursor ions that elute from the high-performance liquid chromatography column. An arbitrary number of runs (stacked mass spectra, left) can be compared using sophisticated alignment and normalization procedures. Quantitative comparison of the isotopic cluster of the same peptide over two runs can be performed. Peptide identities can also be transferred when the peptide is fragmented in only one of the runs but matches precisely the mass and elution time of an aligned peak (known as the ‘match between runs’ feature in MaxQuant¹⁰⁰). Absolute quantities can be estimated by adding up the peak volumes of all peptides that identify a particular protein then determining the proportion of the (known) total proteome mass that has been analysed. Peptides can also be subjected to label-free quantification at the MS² level (right). In this case, the fragment-ion intensities that are unique to a specific peptide are used for quantification, in a way that is analogous to the use of precursor-ion signal intensities for quantification using MS¹-level data. In multiplexed shotgun proteomics, up to ten samples are labelled differentially so that they release reporter ions that can be distinguished in the MS¹ spectra. In DIA-based methods, the intensities of fragments that belong to the same precursor ion are extracted to yield a measure of peptide abundance^{104,105}. Q, quadrupole.

and this has become a crucial requirement for almost all functional studies in the past 10 years.

The preferred method for proteome discovery is data-dependent acquisition (DDA) (Fig. 1) and the past decade has seen striking advances in this area. Whereas the first description of a complete model proteome⁶ and the identification of more than 10,000 different proteins in human cell lines^{7,8} were technological tours de force, a similar depth of coverage can now be achieved within hours and with minimal sample-preparation steps^{9,10}. These developments, although still confined to a few specialized laboratories, will make proteomics increasingly applicable to everyday cell biology and biochemical research, which overwhelmingly uses classic antibody-based techniques such as western blotting. In addition to its exquisite specificity, other advantages of DDA-based proteomics include that it is unbiased and free from hypotheses; that is, the researcher does not need to know the identity of the expected proteins in advance. Furthermore, in a DDA-based proteomics experiment all proteins can be interrogated at once. As well as helping to answer a specific question, proteomics can therefore turn every experiment into a global discovery study, which enables the detection of new and unexpected molecules and connections, providing fresh biological insights. These developments

are supported by publicly accessible bioinformatics tools for processing and interpreting the large amounts of data that are generated in complex projects (Fig. 1). The continued development of highly streamlined and robust proteomics workflows, including robust and economical mass spectrometers, is advocated to usher in an age of complete, accurate and ubiquitous proteomes¹¹, in analogy to what the introduction of next-generation sequencing has provided for genomics-related fields.

Present technology already enables analysis of the complete protein inventory of biological systems, including cell-type-specific proteomes of mammalian organs^{12–14}. One outcome of in-depth proteomics studies has been a demonstration of the extent to which diverse cellular systems have similar proteomes, with few proteins being uniquely detectable in specific situations¹⁵. This surprising finding is supported by the Human Protein Atlas, a large-scale antibody-based study that also reports ubiquitous expression¹⁶. The identity of cells and tissues therefore seems to be determined primarily by the abundance at which they express their constituent proteins, and perhaps by the manner in which the proteins are organized in the proteome, rather than the presence or absence of certain proteins.

The application of DDA-based proteomics to a collection of human

BOX 1

Bottom-up proteomics

Proteins can be studied as intact entities by mass spectrometry, an approach called top-down proteomics²¹. This has the advantage that all modifications that occur on the same molecule can, in principle, be measured together, enabling identification of the precise proteoform¹⁰⁷. However, bottom-up proteomics, in which peptides are generated by the enzymatic digestion of proteins, has been experimentally and computationally more tractable and is the most widespread proteomic workflow. A number of bottom-up techniques exist; each has a specific purpose, a performance profile and a range of utility. In all of the techniques, proteins are extracted from the source material then digested into peptides by a sequence-specific enzyme such as trypsin. The resulting mixture of peptides is separated by reverse-phase chromatography, which is coupled online to electrospray ionization (Fig. 1). The peptide ions are then transferred to the vacuum of a mass spectrometer, where they are fragmented in the gas phase to generate MS/MS (MS^2) spectra that contain the information to identify and quantify specific peptides. Almost always, collision-induced dissociation or higher-energy collisional dissociation¹⁰⁸ are used for fragmentation, but alternative methods are becoming more widely available. One such method, electron transfer dissociation¹⁰⁹, is particularly beneficial for the fragmentation of large and modified peptides. The resulting data are analysed by mass-spectrometry-specific computational pipelines as well as general downstream systems-biology solutions that are tailored to proteomics¹⁰¹.

Three main approaches are used in bottom-up proteomics: discovery (or shotgun) proteomics by means of DDA, aimed at achieving unbiased and complete coverage of the proteome; targeted proteomics using selected reaction monitoring, aimed at the reproducible, sensitive and streamlined acquisition of a subset of known peptides of interest; and multiplexed fragmentation of all peptides that elute from the high-performance liquid chromatography column by DIA, aimed at generating comprehensive fragment-ion maps for a sample (Fig. 1a–c).

In DDA-based methods, mass spectra of all the ion species that co-elute at a specific point in the gradient elution (that is, precursor-ion spectra) are recorded at the MS^1 (or full-scan) level. The instrument alternates between the acquisition of full-scan data and the acquisition of fragment-ion spectra, in which as many precursors as possible are sequentially isolated and fragmented (at the MS^2 level). Of many possible instrument configurations, quadrupole–orbitrap analysers¹¹⁰ dominate DDA proteomics but time-of-flight instruments also have unique promise. In typical ‘top N ’ cycles (in which ‘ N ’ denotes the number of MS^2 spectra that follow), an MS^1 scan is followed by about ten fragment-ion scans. Contemporary instruments transfer ions into the vacuum with greatly improved efficiency, which results in very bright beams (of more than 10^9 ions per second). The resolution of orbitraps has improved several fold, enabling very fast top N cycles

at high resolution. However, the capacity of orbitraps is still limited to about 1 million ions, which restricts the dynamic range that can be achieved in MS^1 spectra.

In targeted proteomics, the proteins of interest are predetermined and known. Using pre-existing information, characteristic (proteotypic) peptides are selectively and recursively isolated and then fragmented over their chromatographic elution time. This is done by setting the first quadrupole of a triple quadrupole instrument to the expected precursor ion m/z ratio and the third quadrupole to the m/z ratio of an abundant fragment ion that is specific for the targeted peptide. (The second quadrupole houses the collision chamber.) To achieve selectivity, the process is multiplexed to several fragments per peptide (known as multiple reaction monitoring, MRM), and throughput is increased by multiplexing it to many peptides¹¹¹. Alongside the robust and economical triple quadrupole instruments, high-resolution instruments such as quadrupole orbitraps are used increasingly for targeted analysis, a variant known as parallel reaction monitoring because it utilizes the entire MS^2 spectrum¹¹².

In DIA-based methods¹¹³ such as SWATH¹⁰³, entire ranges of precursors are fragmented at the same time. The peptide fragmentation information is retrieved from the multiplexed MS^2 spectra either by targeted signal extraction on the basis of previously acquired single-peptide fragmentation spectra¹¹² or by the generation of ‘pseudo’ fragment-ion spectra constructed directly from the DIA data that are then subjected to classic database searching¹⁰⁵. The advantage of this approach is that the entire range of possible precursor-ion masses can be analysed seamlessly and in rapid succession, which eliminates the missing value problem of DDA (in which peptides are only measured in some of a set of liquid chromatography–mass spectrometry (LC– MS^2) runs), at least within the dynamic range that is achieved in the experiment. At present, DIA is limited to a dynamic range of 4–5 orders of magnitude and it requires the *a priori* construction of fragment-ion spectra for the query peptides to deconvolve these peptides from the DIA data^{104,105,114}.

Each of these approaches has advantages and limitations; hybrid methods that combine the best aspects will therefore probably emerge in the near future. Entirely new methods will also be created. For instance, in the past year it has become possible to store several precursor ions in parallel in a trapped-ion mobility device, which can then be followed by serial fragmentation. Known as parallel accumulation–serial fragmentation (PASEF), this method promises to increase the speed and sensitivity of fragmentation several fold¹¹⁵.

Metabolic and chemical labelling strategies have matured and can now be used for precise quantification, but they can still suffer from limitations to their accuracy and dynamic range^{116–118}. Improvements in the resolution that can be achieved, combined with advances in algorithms, are making label-free quantification increasingly useful for DDA¹¹⁹, selected reaction monitoring¹²⁰ and DIA^{104,105} methods.

tissues, combined with the integration of data from the community, has resulted in two draft human proteomes^{17,18}. Mass-spectrometric evidence for 84% (ref. 17) or 92% (ref. 18) of protein-coding sequences was reported. However, re-analysis of the data using standard and community-approved false-discovery rates for peptides and proteins leads to much lower coverage and the removal of proteins not thought to be expressed in the sampled tissues^{19,20}. Extensive peptide pre-fractionation has been combined with digestion by various enzymes and peptide fragmentation methods to reach a depth of proteome coverage that should soon be on par with the comprehensiveness to which the transcriptome can be probed by next-generation sequencing¹³. Comprehensive

characterization of the proteome is therefore feasible and we predict that it will soon become routine¹¹. The coverage of identified proteins with sequenced peptides has also been improving, which makes it increasingly realistic to distinguish between and quantify proteoforms, the different molecular forms of a protein that originate from the same gene. A complete inventory of proteoforms cannot yet be achieved and will be a challenge to attain because of the combinatorial explosion of proteoforms that are created by even a moderate number of modifications. Top-down proteomics characterizes the actual combination of modification events for each proteoform²¹. Although attractive in principle, top-down mass spectrometry is experimentally and computationally challenging because

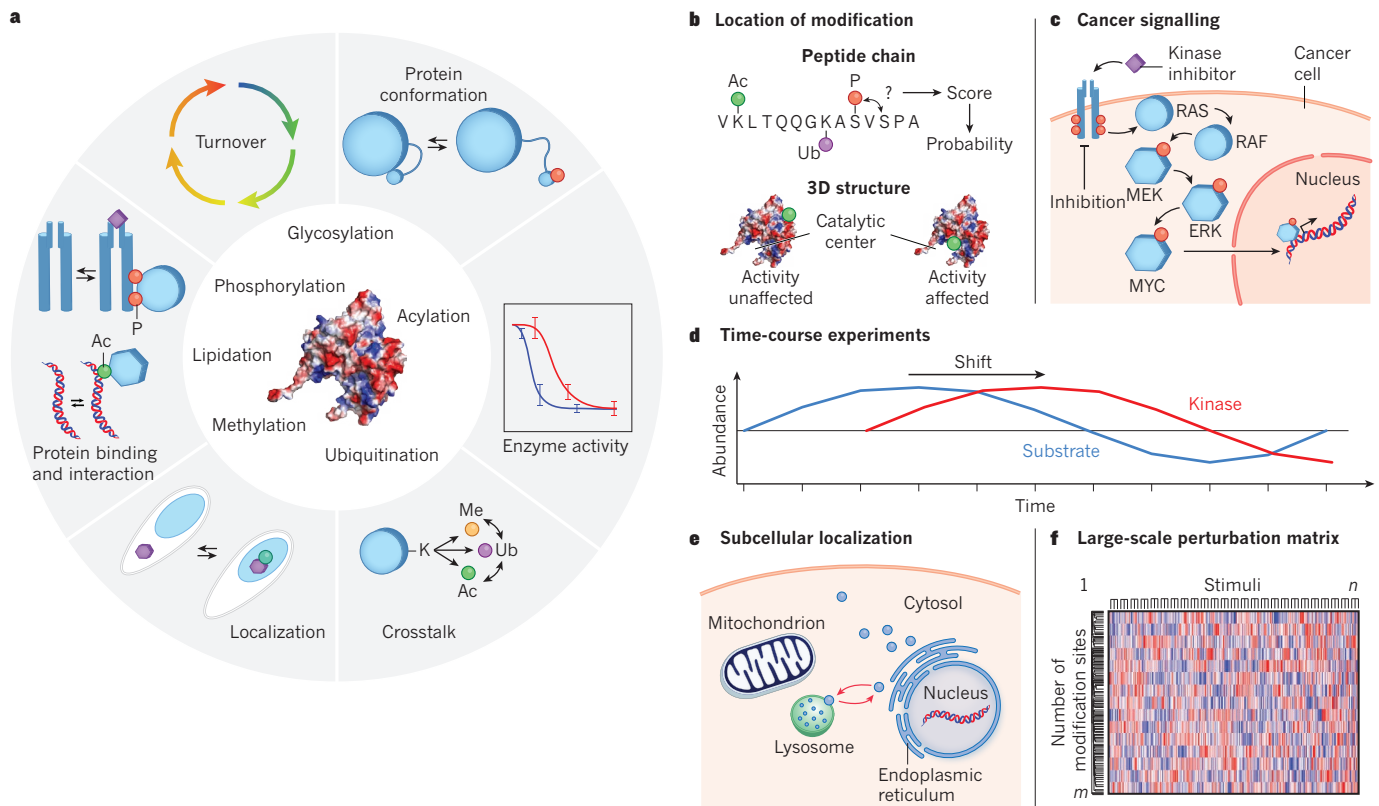


Figure 2 | Analysis of post-translational modifications. **a**, In post-translational modification, proteins are modified through the attachment of a chemical moiety such as a phosphate group, usually by a dedicated and highly specific system of enzymes. The most commonly studied post-translational modifications are listed (centre) and these are accompanied by hundreds of other less-well-studied or unknown types of modifications. Such modifications can lead to: alterations in protein conformation (through phosphorylation) and subsequent allosteric regulation; changes in enzyme activity; crosstalk that results from the same amino-acid residue being targeted by more than one type of modification; alterations in the subcellular localization of proteins; changes in protein binding; and alterations in protein lifetimes (for example, through the attachment of the small protein ubiquitin). Ac, acetyl; ERK, extracellular signal-related kinase; Me, methyl; MEK, mitogen-activated protein kinase kinase; MYC, transcription factor cMYC; P, phosphate; RAF, RAF kinase; RAS, RAS GTPase; Ub, ubiquitin. **b**, After a modified peptide has been identified from the fragment spectra, the amino acid in the peptide chain to which the post-translational modification is attached must be determined. The location of the modification within the three-dimensional structure of the protein can

of the greater difficulty in analysing proteins in comparison with peptides and because each protein is distributed as multiple proteoforms that might or might not differ functionally. The array of modern mass spectrometry techniques has also been deployed to analyse unique types of sample with biological and clinical importance, including secreted proteins in the context of immunology²², the peptidome of body fluids such as cerebrospinal fluid²³, the immunopeptidome²⁴ and the extracellular matrix²⁵.

Proteomics is sufficiently advanced to warrant the in-depth characterization of a great variety of biological systems. Along with other important information, this enables protein copy numbers or concentrations to be determined on a proteome-wide scale^{26–28}, which helps to improve understanding of the underlying biology.

Characterizing protein modifications and cell signalling

Mass-spectrometry-based proteomics is well suited to the study of post-translational modifications because such changes lead to characteristic shifts in mass and can be located with the resolution of a single amino acid through peptide-fragment ion spectra (Fig. 2). The only deviation from

often also be determined, which provides clues about function. **c**, Global interrogation of the changes in a signalling pathway can be achieved readily by quantitative phosphoproteomics. For example, the suppression of aberrant signalling in cancer cells by drugs known as kinase inhibitors can be followed. **d**, Detailed time-course experiments yield information on the temporal ordering of events such as the activation of a kinase upstream of one of its substrates. The proportion of proteins that are modified by a particular post-translational modification (also termed the occupancy or stoichiometry) can change drastically depending on the biological conditions (not shown). It can be derived from the changes in protein level and the levels of the modified and unmodified peptide in two cellular states¹⁰⁶. **e**, The modification of a protein often determines its subcellular localization — that is, whether it is found in the nucleus or the cytosol, for instance. Many types of stimuli can be applied to biological systems, after which the level of a particular post-translational modification can be determined. **f**, The structure of the perturbation matrix that results reveals the regulated sites and how they correlate between stimuli, as indicated by hot spots in the heat map. *m*, number of modification sites quantified; *n*, number of stimuli applied.

the DDA-based proteomic workflow that is used to identify unmodified peptides is the addition of an enrichment step for peptides that carry the modification of interest. Post-translational modifications that are particularly labile, such as *O*-linked β -*N*-acetylglucosamine (*O*-GlcNAc), benefit from the use of electron transfer dissociation as the fragmentation method, and certain classes of modifications, including glycosylations with large glycans and nucleotide modifications, can also be challenging to detect using mass spectrometry. The most frequently studied types of post-translational modifications are phosphorylation, ubiquitylation, the addition of ubiquitin-like proteins, glycosylation, methylation, acetylation and other types of acylation. For these, present technology enables the identification of thousands of sites of modification and their accurate quantification between proteomic states²⁹. The main surprise has been the number and diversity of these post-translational modifications as well as how many of them seem to be involved in cellular regulation. For example, more than 50,000 phosphorylation events on at least 75% of the proteome have been documented in a single cell line³⁰. Phosphoproteomics is used routinely to quantify the response of cells to

stimuli and such studies have reached a remarkable level of detail and sophistication. As well as providing large catalogues of sites, they have led to the discovery of sites of regulation with pivotal roles in determining the state of biological processes^{31–35}. A streamlined protocol has made it possible to analyse *in vivo* signalling events with high temporal resolution³⁶. This revealed that insulin signalling in the liver is unexpectedly fast: maximal phosphorylation was reached within a few seconds at many sites and transcription factors were phosphorylated fully within 30 seconds. Another message emerging from phosphoproteomics is that the proportion of sites that are functional seems to be high. This is suggested by high stoichiometry (that is, the fraction of proteins that are phosphorylated at a specific site), a large number of highly regulated sites in diverse processes, and by the tight temporal correlation of many uncharacterized sites with sites that are known to be functional. Conversely, lysine acetylation behaves very differently: the stoichiometry is extremely low for most sites and often these modifications seem to be of a non-enzymatic origin, which is also true for acylations such as succinylation^{37,38}. Lysine is the most frequently modified amino-acid residue and the specific target of ubiquitylation, a modification that can be enriched efficiently and studied in a linkage-specific manner by mass spectrometry. Effective strategies also exist for characterizing SUMOylation and modification with other ubiquitin-like proteins, and these have revealed unique insights into their large-scale behaviour³⁹. Histone modifications and their regulators (proteins known as ‘writers’, ‘readers’ and ‘erasers’ that make, recognize and edit epigenetic marks) are of great interest and specific methods have been devised for their detection^{40,41}.

Mass spectrometry also enables the characterization of hundreds of exotic or unknown modifications^{42–44}. This emerging area builds on new instrumentation, innovative methods of fragmentation and fresh protocols for enrichment but faces the challenge of devising enrichment methods that are specific for each post-translational modification of interest. As the proteome is probed to ever increasing depths, the analysis of modifications without their enrichment is becoming more feasible, and this is already possible for methylation and phosphorylation.

Post-translational modifications and proteolytic processing events, in particular, can also be analysed using chemical proteomics approaches. These use compounds that bind to engineered small-molecule binding pockets⁴⁵ or probes that label the freshly created N termini of proteins after

cleavage^{46,47}. The deep, quantitative and time-resolved analysis of specific types of modifications in many systems and species has already provided a wealth of biological insights. These data also indicate that specific modification systems intersect and cooperate to generate a specific cellular state. The comprehensive analysis of proteoforms that differ in their state of modification, the determination of the functional significance of such proteoforms and the elucidation of the processes that catalyse and control their homeostasis remain challenges for the future.

Protein modules, networks and cellular functions

Proteins rarely function alone; instead, they depend on the association of various components into macromolecular complexes. The concept of modular biology, proposed by Leland Hartwell and his colleagues, states that the biological functions of the cell are carried out by multicomponent modules⁴⁸, and the modularity of the proteome has been impressively demonstrated by several classic studies⁴⁹. An array of mass-spectrometry-based strategies, the best established of which is interaction proteomics, has made considerable contributions to integrative or hybrid approaches to yield the composition, topology and structure of specific complex macromolecular assemblies⁵⁰.

Interaction proteomics involves a pull-down assay of a bait protein with its binding partners followed by mass-spectrometric analysis, known as affinity-purification mass spectrometry (AP-MS)⁵¹ (Fig. 3a). Thousands of proteins can be detected in such experiments owing to the high sensitivity of mass spectrometry and the propensity of the samples to contain unspecific contaminants. Proteins that bind with specificity to the bait can be distinguished effectively from the contaminants through the quantitative comparison of samples with control assays, preferably using rigorous statistical controls^{52,53}. Without the ability to distinguish background binding, the reported interactomes of specific proteins often contain hundreds of purported binders with little biological importance. Versions of this basic AP-MS workflow have been implemented robustly to support large-scale mapping of the wiring diagrams of the human cellular proteome⁵⁴. Taking advantage of the relative abundance levels of prey proteins and the endogenously expressed bait, and adding copy numbers of the entire cellular proteome, provides a human interactome in three quantitative dimensions and enables the estimation of binding stoichiometries. This helps to classify interactions into stable, regulatory

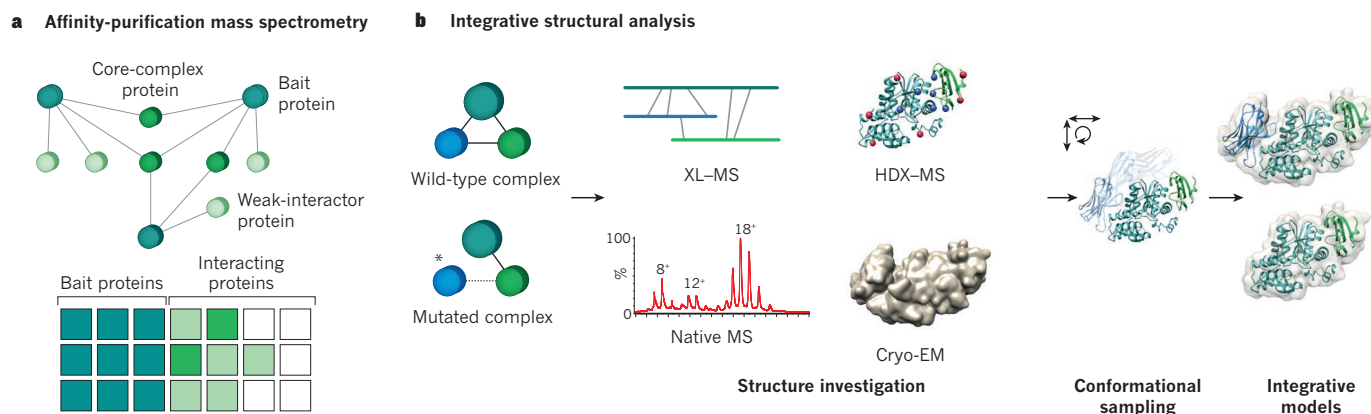


Figure 3 | Interaction proteomics and structural proteomics. **a**, Schematic representations of a protein interaction network with bait proteins (teal), core complex members (dark green) and weak interactors (light green). A bait protein is precipitated with its interaction partners and is measured in replicates by one of the workflows described in Fig. 1. By considering the interaction stoichiometry (the molar ratio of prey proteins and the bait protein expressed under endogenous control) and the relative cellular abundances of the proteins, stable core complexes can be distinguished from weak interactions and unspecific interactions, as well as from asymmetric interactions between proteins of different abundances⁵⁵. **b**, A wild-type protein complex and the same complex with mutations (*) are investigated using complementary structural

techniques, collectively termed integrative or hybrid structural analysis. For example, XL-MS can reveal information about subunit topology and direct domain–domain interactions. Hydrogen–deuterium exchange mass spectrometry (HDX-MS) is able to determine the interaction surfaces and solvent-exposed regions. Native mass spectrometry (native MS), in which entire protein complexes are electrosprayed into the mass spectrometer, can infer the stoichiometry and the assembly pathway of such complexes, and cryo-EM can obtain their overall shape and their density maps. The heterogeneous structural restraints are integrated in a common computational framework that evaluates subunit configurations (known as conformational sampling). Consensus models that represent the structures of the wild-type and mutated complexes can then be derived.

or transient ones and even captures client interactions such as proteins being folded by chaperone complexes⁵⁵. This work established that networks of cells are surprisingly dominated by a large number of weak interactions and that the number of stable core complexes is limited. The emerging picture of a modular proteome in which modules have variable stoichiometric robustness is also supported by a study in which the relative changes of bona fide protein components of 182 complexes were determined in 11 cell types and 5 temporal states⁵⁶. The covariance of the co-expression profiles for complex subunits varied considerably, which suggests that dynamic subunit associations fine-tune the composition and function of specific cellular modules⁵⁶.

Modified peptides, oligonucleotides and small molecules have also been used with success as bait proteins for AP-MS experiments⁵¹. For instance, transcription-factor complexes that are crosslinked to DNA can be analysed readily, as can protein complexes that are recruited to specific DNA lesions⁵⁷. Other approaches to capture protein interactions include enzyme-mediated proximity labelling in cells followed by pull-down assays of the labelled proteins^{58,59} and the accurate measurement of co-fractionation patterns^{60–62}. Such measurements are also the basis of organellar proteomics, which aims to determine the subcellular location and dynamics of the proteome^{63–66}, a valuable complement to imaging-based technologies.

Although AP-MS and related methods indicate the composite population of proteins that is associated with a particular bait protein, other mass-spectrometry-based methods can also identify the subunit interfaces, topology, conformation and structure of protein complexes (Fig. 3b), as shown by the analysis of the nuclear pore complex⁶⁷.

Native mass spectrometry, which is the direct analysis of macromolecular assemblies by mass spectrometry, has been used both by itself⁶⁸ and as part of an integrative approach⁶⁹ to gain insights into the subunit stoichiometry, topology and structure of macromolecular assemblies. When applied to membrane protein complexes, the technique revealed an unappreciated structural role for lipids in respiratory protein complexes⁷⁰.

Integrative or hybrid approaches complement X-ray crystallography and nuclear magnetic resonance, methods that are central to structural biology, and mass spectrometry has become an essential component of the hybrid structural-biology toolbox⁷¹. Distance restraints that are generated by chemical crosslinking and the mass-spectrometry-based identification of crosslinked residues (an approach termed XL-MS) have proven helpful for determining the structure of large complexes⁷², particularly in combination with single-particle cryo-electron microscopy (cryo-EM) data. XL-MS and cryo-EM have been used to solve longstanding problems in structural biology⁷¹, to identify the substrate binding sites in molecular chaperones⁷³ and to detect steric alterations in complexes in different functional states⁷⁴. XL-MS has also been used to analyse protein-RNA interfaces⁷⁵, to identify receptor-ligand pairs directly⁷⁶, to map physical interactions between different types of biomolecules and to identify the ligands of orphan receptors.

Integrative structural-biology methods are being adapted for use with the microgram amounts of protein complexes that are isolated by affinity purification, and this advance has been applied to mapping the organization of the protein phosphatase 2A (PP2A) enzyme system in HEK293 cells⁷⁷. Using the two catalytic subunits, the scaffold subunit and most of the 15 regulatory subunits from which trimeric PP2A structures are combinatorially assembled as bait proteins, XL-MS identified the protein-protein interfaces, the actual subunit composition of the PP2A complexes that are concurrently expressed in the cell and their associated proteins to establish a high-granularity protein interaction network consisting of more than 150 proteins⁷⁷.

Notably, XL-MS is beginning to be used on a proteomics scale^{78,79}. Although the crosslinks that are identified in such studies come primarily from highly expressed complexes, they highlight a path towards the direct measurement of protein-protein interfaces in the cell. The combination of AP-MS and XL-MS was recently refined so that chemical crosslinks could be identified from samples containing only a few million cells^{80,81}. Complexes that are isolated by AP-MS can also be used to

generate cryo-EM single-particle data, which opens up the possibility of linking the atomic structure and function of macromolecular assemblies that have been isolated from cells in a particular functional state. Results from cryo-electron tomography studies further extend this perspective towards the possibility of observing specific macromolecular modules by template matching *in situ*^{82,83}.

In a similar way to their composition, the conformation of the subunits of protein complexes can adapt to the state of the cell. Mass spectrometry techniques can detect changes in protein conformation and protein interfaces and then relate these observations to functional alterations in particular proteins. Hydrogen-deuterium exchange mass spectrometry is a classic method for determining alterations in the conformation, structure and interfaces of specific complexes⁸⁴. By contrast, the hydroxyl radical footprinting method predominantly labels solvent-exposed side chains and is not affected by back exchange of the labelled residues⁸⁵. The different conformations of a protein can vary in thermal stability, an observation that has been used to probe conformational changes at a proteomic scale⁸⁶. Cells treated with a cancer drug were subjected to different temperatures, after which heat-denatured proteins were removed and the remaining soluble proteins were analysed by mass spectrometry. This pinpointed both expected and unexpected binding partners of the drug. A conceptually similar technique used the fact that conformational changes in proteins can be detected using protein digestion patterns generated under conditions of limited proteolysis⁸⁷. Structural features of more than 1,000 yeast proteins were concurrently monitored by targeted mass spectrometry and altered conformations for about 300 proteins on a change in nutrients were detected⁸⁷. Such examples demonstrate how structural proteomics techniques are helping to tackle the challenge of detecting often weak interactions between proteins, small-molecule ligands and cofactors on a global scale, as well as the structural effects of ligand binding.

Proteotype states and cellular phenotypes

In the 1940s, Linus Pauling established that a structural alteration in haemoglobin was related causally to a disease phenotype⁸⁸. In that particular case, the structural variation was caused by a single amino acid change in one of the haemoglobin chains, the result of a mutation in the gene that encodes the chain. The extension of this fundamental principle of biology to the level of proteome networks suggests that genetic or external perturbations change the state of the proteome network and that such changes cause or correlate with altered phenotypes (Fig. 4). The state of a proteome that is associated with a specific phenotype can be described as a proteotype. The association between a proteotype and its corresponding phenotype can be investigated by means of two mass-spectrometry-based approaches that differ in principle. The first approach attempts to describe a phenotype mechanistically using the aggregated structure and function of the proteins or modules that constitute the underlying processes. The second approach associates a phenotype with its proteotype through advanced statistical machine-learning tools (known collectively as ‘big data’ analytics) but does not necessarily reach a causal or mechanistic understanding of the underlying processes. Both approaches have been greatly advanced by mass-spectrometry-based technology. In particular, the big data approach based on statistical associations has become possible only through the development of mass spectrometry techniques that are capable of quantifying sets of proteins with a high degree of reproducibility across large collections of samples, generating large data matrices of proteins measured across various samples with minimal missing values. Mass-spectrometry techniques that are used to generate such matrices include the matching of MS¹ intensity maps, using their retention time versus mass-to-charge ratio, from collections of samples and DIA-based methods, and the targeted mass spectrometry of smaller numbers of proteins (Box 1).

In a demonstration of these concepts, a yeast genetic reference panel was used to quantify the effect of genetic perturbations on a metabolic network⁸⁹. Selected-reaction-monitoring targeted mass spectrometry measured 50 metabolic proteins in 96 genetically well-defined strains of yeast. Parental strains acquired independent genetic variations that

consistently affected levels of proteins from the same module or pathway that selective pressures favoured for the acquisition of sets of polymorphisms that maintain the stoichiometry of complexes and pathways. Similarly, 192 proteins that constituted a metabolic network were quantified by selected reaction monitoring of liver samples in two metabolic states from 40 strains of mice from a genetic reference strain compendium⁹⁰, enabling genetic and environmental perturbations to be probed effectively⁹¹. This established a direct mechanistic link between alleles of the gene *Dhtkd1* (a protein quantitative trait locus (pQTL)), the quantity of 2-aminoadipate (a metabolite that is controlled by *Dhtkd1*) and a disease risk for type 2 diabetes. Mechanistic and data-driven approaches can therefore converge to enhance understanding of complex phenotypes if multilevel omics data are integrated at the level of modular networks. Repeating the proteomics measurements of the liver samples using DIA-based mass spectrometry techniques quantified more than 2,600 proteins across the collection of samples, which led to the detection of hundreds of pQTLs as well as mechanistic insights into inborn errors of metabolism and the determination of a molecular basis for respiratory super-complex formation⁹².

These examples and analogous ones from the proteogenomics of cancer⁹³ establish a link through association studies between genetic loci and the network state, as well as between the network state and disease phenotypes. The mass spectrometry methods of bottom-up proteomics (Box 1) represent a general experimental framework for systematically probing the proteotype at ever increasing levels of completeness and precision to support the association of proteotypes and phenotypes.

In the context of translational medicine, proteins that consistently alter their abundance in correlation with a disease phenotype are considered to be biomarker candidates for the phenotype of interest. Typically, a small number of study participants are investigated in depth to extract potential biomarkers that can be validated in larger cohorts^{94,95}. Although attractive in principle, biomarker discovery using mass-spectrometry-based methods is extremely challenging in practice. However, data-driven approaches are opening fresh avenues to associating protein-expression patterns with disease states.

In particular, the detection of protein biomarkers in blood plasma as a window to the physiological state of a person has been an important goal of protein science since before the advent of mass spectrometry. Experience gained over the past decade in plasma proteome analysis by mass spectrometry has demonstrated the enormous challenges of this approach, which are rooted in the complexity of the plasma proteome, its inherent variability across a population and the prevalence of factors that affect its composition, including age, gender and lifestyle. However, several studies^{94,96,97} have shown that the highly reproducible mass spectrometry techniques used for proteotype measurements in tissues can be applied to plasma proteins. Fast and reliable measurements of plasma samples will therefore be possible in collections that consist of hundreds of samples. The systematic measurement of plasma proteins in twin populations has already been used to associate observed changes in abundance in the plasma proteome with genotype⁹⁸. Furthermore, the plasma proteome can now be probed in a broad and high-throughput manner with the aim of extracting as much information about the health or disease state of an individual as possible, effectively enabling high-throughput phenotyping of people⁹⁶. Continuing advances in mass spectrometry technology might therefore enable the future discovery of clinically actionable protein biomarker patterns.

Outlook

Over the past decade, mass-spectrometry-based proteomics has matured from a largely technology-driven field of research into a mainstream analytical tool for the life sciences. It is a versatile approach that supports the analysis of many aspects of proteins, including sequence, quantity, state of modification, structure and macromolecular context. It also accommodates a variety of research approaches, such as mechanism-oriented exploration for determining causal relationships and big-data strategies that rely on statistical associations to discover biological relationships.

Further, dramatic improvements in the core technology of mass

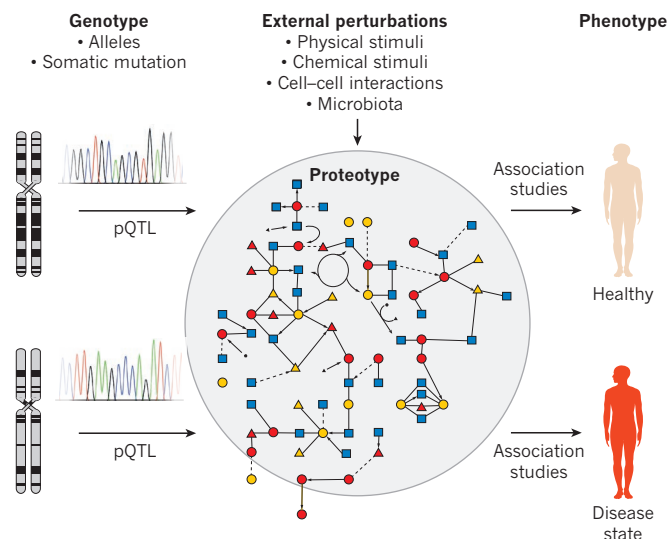


Figure 4 | Proteotype states and phenotypes. The proteotype, which is the acute state of the proteome, is shown as a modular network of interacting protein entities (coloured shapes). The composition of the proteotype and the organization of individual proteins into functional modules and interaction networks are determined by the combined effects of genotype and external perturbations, which include physical or chemical stimuli, cell-cell interactions or the microbiota. Genotypic differences such as allele differences or somatic mutations might perturb the proteotype. The relationship between genetic loci and the abundance of a protein can be described by a pQTL. These are identified by associating the abundance of a specific protein with particular alleles in genetically characterized sample populations such as genetic reference panels. In turn, the proteotype determines phenotypes, including clinical phenotypes. Association studies can identify relationships between proteotypes and phenotypes. Establishing such associations requires the generation of quantitatively accurate and highly reproducible datasets in which the same proteins are quantified across a large number of samples (for example, genetic reference panels or cohorts of patients). Datasets that support such association studies can now be generated using various mass spectrometry techniques.

spectrometry are probable and will open up the field of proteomics to even more applications. Aside from a focus on signalling and structural applications, important goals for proteomics will be to build comprehensive and quantitative catalogues of proteins under many conditions and perturbations and to organize these proteoforms into a modular proteome of the cell. This will improve understanding of processes across many areas of biology and diseases and will constitute an excellent starting point for modelling the cell. For this to occur, proteomics must be tightly integrated with other technologies and it should address challenges such as single-cell analysis, an approach that was pioneered by mass cytometry⁹⁹. The integration of different types of data is already far advanced in the case of next-generation sequencing technologies (for example, RNA sequencing, chromatin immunoprecipitation followed by sequencing (ChIP-seq) and ribosome profiling) and metabolomics, and the integration of data from structural biology and imaging-based technologies is advancing at a rapid pace. There are also considerable opportunities for bringing proteomics together with increasingly efficient tools for editing the genome — in particular, CRISPR-Cas9. We envision this to work in an iterative manner in which proteomics findings are interrogated by deleting, tagging and point-mutating one or more genes of importance, followed by further rounds of proteomics measurements to determine the effects of the genetic alterations on the proteome. This will address the fundamental question of how genotypic variability is mechanistically translated into phenotypic variability. The integration of various omics approaches and many perturbations will generate exponential flows of disparate data types. This will necessitate commensurate advances in bioinformatics and computational proteomics, which will be powered increasingly by

machine-learning technologies while retaining their ability to generate biological insights. In this regard, the journey from single-protein analysis to a true understanding of the proteome and the importance of proteo-types will be long, challenging and exciting. ■

Received 11 January; accepted 15 July 2016.

1. Marguerat, S. *et al.* Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. *Cell* **151**, 671–683 (2012).
2. Milo, R. What is the total number of protein molecules per cell volume? A call to rethink some published values. *BioEssays* **35**, 1050–1055 (2013).
3. Edwards, A. M. *et al.* Too many roads not taken. *Nature* **470**, 163–165 (2011).
4. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).
5. Cravatt, B. F., Simon, G. M. & Yates, J. R. The biological impact of mass-spectrometry-based proteomics. *Nature* **450**, 991–1000 (2007).
6. de Godoy, L. M. F. *et al.* Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455**, 1251–1254 (2008).
This paper demonstrates that complete proteomes of a model organism can be obtained and quantified in different biological states.
7. Beck, M. *et al.* The quantitative proteome of a human cell line. *Mol. Syst. Biol.* **7**, 549 (2011).
8. Nagaraj, N. *et al.* Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* **7**, 548 (2011).
9. Hebert, A. S. *et al.* The one hour yeast proteome. *Mol. Cell. Proteomics* **13**, 339–347 (2014).
10. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nature Methods* **11**, 319–324 (2014).
11. Mann, M., Kulak, N. A., Nagaraj, N. & Cox, J. The coming age of complete, accurate, and ubiquitous proteomes. *Mol. Cell* **49**, 583–590 (2013).
12. Azimifar, S. B., Nagaraj, N., Cox, J. & Mann, M. Cell-type-resolved quantitative proteomics of murine liver. *Cell Metab.* **20**, 1076–1087 (2014).
13. Richards, A. L., Merrill, A. E. & Coon, J. J. Proteome sequencing goes deep. *Curr. Opin. Chem. Biol.* **24**, 11–17 (2015).
14. Sharma, K. *et al.* Cell type- and brain region-resolved mouse brain proteome. *Nature Neurosci.* **18**, 1819–1831 (2015).
15. Lundberg, E. *et al.* Defining the transcriptome and proteome in three functionally different human cell lines. *Mol. Syst. Biol.* **6**, 450 (2010).
16. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
This paper provides an integrative analysis of the human proteome through large-scale antibody localization and transcriptomics; the findings are organized in an accompanying database.
17. Kim, M.-S. *et al.* A draft map of the human proteome. *Nature* **509**, 575–581 (2014).
18. Wilhelm, M. *et al.* Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582–587 (2014).
This study aggregates data on diverse human proteomes from the authors and the research community and, like ref. 17, argues that a large part of the genome is accessible to mass-spectrometric detection.
19. Ezkurdia, I., Vázquez, J., Valencia, A. & Tress, M. Analyzing the first drafts of the human proteome. *J. Proteome Res.* **13**, 3854–3855 (2014).
20. Omenn, G. S. *et al.* Metrics for the Human Proteome Project 2015: progress on the human proteome and guidelines for high-confidence protein identification. *J. Proteome Res.* **14**, 3452–3460 (2015).
21. Tran, J. C. *et al.* Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **480**, 254–258 (2011).
22. Meissner, F., Scheltema, R. A., Mollenkopf, H.-J. & Mann, M. Direct proteomic quantification of the secretome of activated immune cells. *Science* **340**, 475–478 (2013).
23. Secher, A. *et al.* Analytic framework for peptidomics applied to large-scale neuropeptide identification. *Nature Commun.* **7**, 11436 (2016).
24. Caron, E. *et al.* Analysis of major histocompatibility complex (MHC) immunopeptides using mass spectrometry. *Mol. Cell. Proteomics* **14**, 3105–3117 (2015).
25. Schiller, H. B. *et al.* Time- and compartment-resolved proteome profiling of the extracellular niche in lung injury and repair. *Mol. Syst. Biol.* **11**, 819 (2015).
26. Malmström, J. *et al.* Proteome-wide cellular protein concentrations of the human pathogen *Leptospira interrogans*. *Nature* **460**, 762–765 (2009).
27. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
A pioneering investigation of the degree of correlation between the transcriptome and the proteome — a question that is still unresolved.
28. Wisniewski, J. R., Hein, M. Y., Cox, J. & Mann, M. A “proteomic ruler” for protein copy number and concentration estimation without spike-in standards. *Mol. Cell. Proteomics* **13**, 3497–3506 (2014).
29. Doll, S. & Burlingame, A. L. Mass spectrometry-based detection and assignment of protein posttranslational modifications. *ACS Chem. Biol.* **10**, 63–71 (2015).
30. Sharma, K. *et al.* Ultra-deep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. *Cell Rep.* **8**, 1583–1594 (2014).
31. Hsu, P. P. *et al.* The mTOR-regulated phosphoproteome reveals a mechanism of mTORC1-mediated inhibition of growth factor signaling. *Science* **332**, 1317–1322 (2011).
32. Huttlin, E. L. *et al.* A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143**, 1174–1189 (2010).
33. Olsen, J. V. *et al.* Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635–648 (2006).
34. Francavilla, C. *et al.* Functional proteomics defines the molecular switch underlying FGF receptor trafficking and cellular outputs. *Mol. Cell* **51**, 707–722 (2013).
35. Steger, M. *et al.* Phosphoproteomics reveals that Parkinson’s disease kinase LRRK2 regulates a subset of Rab GTPases. *eLife* **5**, e12813 (2016).
This study used a combination of genetics, chemical proteomics and cutting-edge phosphoproteomics to reveal genuine, *in vivo* substrates of the Parkinson’s disease kinase LRRK2, opening the way to clinical trials.
36. Humphrey, S. J., Azimifar, S. B. & Mann, M. High-throughput phosphoproteomics reveals *in vivo* insulin signaling dynamics. *Nature Biotechnol.* **33**, 990–995 (2015).
37. Weinert, B. T. *et al.* Acetyl-phosphate is a critical determinant of lysine acetylation in *E. coli*. *Mol. Cell* **51**, 265–272 (2013).
38. Choudhary, C., Weinert, B. T., Nishida, Y., Verdin, E. & Mann, M. The growing landscape of lysine acetylation links metabolism and cell signalling. *Nature Rev. Mol. Cell Biol.* **15**, 536–550 (2014).
39. Hendriks, I. A. *et al.* Uncovering global SUMOylation signaling networks in a site-specific manner. *Nature Struct. Mol. Biol.* **21**, 927–936 (2014).
40. Huang, H., Lin, S., Garcia, B. A. & Zhao, Y. Quantitative proteomic analysis of histone modifications. *Chem. Rev.* **115**, 2376–2418 (2015).
41. Zheng, Y., Huang, X. & Kelleher, N. L. Epiproteomics: quantitative analysis of histone marks and codes by mass spectrometry. *Curr. Opin. Chem. Biol.* **33**, 142–150 (2016).
42. Savitski, M. M., Nielsen, M. L. & Zubarev, R. A. ModifiComb, a new proteomic tool for mapping substoichiometric post-translational modifications, finding novel types of modifications, and fingerprinting complex protein mixtures. *Mol. Cell. Proteomics* **5**, 935–948 (2006).
43. Jungmichel, S. *et al.* Proteome-wide identification of poly(ADP-ribosyl)ation targets in different genotoxic stress responses. *Mol. Cell* **52**, 272–285 (2013).
44. Chick, J. M. *et al.* A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides. *Nature Biotechnol.* **33**, 743–749 (2015).
45. Rix, U. & Superti-Furga, G. Target profiling of small molecules by chemical proteomics. *Nature Chem. Biol.* **5**, 616–624 (2009).
46. Gawron, D., Ndah, E., Gevaert, K. & Van Damme, P. Positional proteomics reveals differences in N-terminal proteoform stability. *Mol. Syst. Biol.* **12**, 858 (2016).
47. Kleifeld, O. *et al.* Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. *Nature Protocols* **6**, 1578–1611 (2011).
48. Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402** (suppl.), C47–C52 (1999).
49. Pawson, T. Protein modules and signalling networks. *Nature* **373**, 573–580 (1995).
50. Ward, A. B., Sali, A. & Wilson, I. A. Integrative structural biology. *Science* **339**, 913–915 (2013).
51. Dunham, W. H., Mullin, M. & Gingras, A.-C. Affinity-purification coupled to mass spectrometry: basic principles and strategies. *Proteomics* **12**, 1576–1590 (2012).
52. Choi, H. *et al.* SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nature Methods* **8**, 70–73 (2011).
53. Keilhauer, E. C., Hein, M. Y. & Mann, M. Accurate protein complex retrieval by affinity enrichment mass spectrometry (AE-MS) rather than affinity purification mass spectrometry (AP-MS). *Mol. Cell. Proteomics* **14**, 120–135 (2015).
54. Huttlin, E. L. *et al.* The BioPlex network: a systematic exploration of the human interactome. *Cell* **162**, 425–440 (2015).
A large-scale investigation of proteins binding to tagged constructs to establish a human interactome.
55. Hein, M. Y. *et al.* A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* **163**, 712–723 (2015).
This paper describes the characterization of a human interactome using bait proteins that are expressed under endogenous control; its analysis in several quantitative dimensions revealed a preponderance of weak interactions.
56. Ori, A. *et al.* Spatiotemporal variation of mammalian protein complex stoichiometries. *Genome Biol.* **17**, 47 (2016).
57. Räsche, M. *et al.* Proteomics reveals dynamic assembly of repair complexes during bypass of DNA cross-links. *Science* **348**, 1253671 (2015).
58. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* **196**, 801–810 (2012).
59. Rhee, H.-W. *et al.* Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science* **339**, 1328–1331 (2013).
60. Havugimana, P. C. *et al.* A census of human soluble protein complexes. *Cell* **150**, 1068–1081 (2012).
61. Kristensen, A. R., Gsponer, J. & Foster, L. J. A high-throughput approach for measuring temporal changes in the interactome. *Nature Methods* **9**, 907–909 (2012).
62. Wan, C. *et al.* Panorama of ancient metazoan macromolecular complexes. *Nature* **525**, 339–344 (2015).
63. Christoforou, A. *et al.* A draft map of the mouse pluripotent stem cell spatial proteome. *Nature Commun.* **7**, 8992 (2016).
64. Larance, M. & Lamond, A. I. Multidimensional proteomics for cell biology. *Nature Rev. Mol. Cell Biol.* **16**, 269–280 (2015).

65. Yates, J. R., Gilchrist, A., Howell, K. E. & Bergeron, J. J. M. Proteomics of organelles and large cellular structures. *Nature Rev. Mol. Cell Biol.* **6**, 702–714 (2005).
66. Itzhak, D. N., Tyanova, S., Cox, J. & Borner, G. H. Global, quantitative and dynamic mapping of protein subcellular localization. *eLife* **5**, e16950 (2016).
67. Alber, F. *et al.* The molecular architecture of the nuclear pore complex. *Nature* **450**, 695–701 (2007).
68. Marcoux, J. & Robinson, C. V. Twenty years of gas phase structural biology. *Structure* **21**, 1541–1550 (2013).
69. Politis, A. *et al.* A mass spectrometry-based hybrid method for structural modeling of protein complexes. *Nature Methods* **11**, 403–406 (2014).
70. Zhou, M. *et al.* Mass spectrometry of intact V-type ATPases reveals bound lipids and the effects of nucleotide binding. *Science* **334**, 380–385 (2011).
- An elegant demonstration of native mass spectrometry in structural studies of intact membrane complexes.**
71. Leitner, A., Faini, M., Stengel, F. & Aebersold, R. Crosslinking and mass spectrometry: an integrated technology to understand the structure and function of molecular machines. *Trends Biochem. Sci.* **41**, 20–32 (2016).
72. Liu, F. & Heck, A. J. Interrogating the architecture of protein assemblies and protein interaction networks by cross-linking mass spectrometry. *Curr. Opin. Struct. Biol.* **35**, 100–108 (2015).
73. Joachimiak, L. A., Walzthoeni, T., Liu, C. W., Aebersold, R. & Frydman, J. The structural basis of substrate recognition by the eukaryotic chaperonin TRiC/CCT. *Cell* **159**, 1042–1055 (2014).
74. Walzthoeni, T. *et al.* xTract: software for characterizing conformational changes of protein complexes by quantitative cross-linking mass spectrometry. *Nature Methods* **12**, 1185–1190 (2015).
75. Kramer, K. *et al.* Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nature Methods* **11**, 1064–1070 (2014).
76. Frei, A. P. *et al.* Direct identification of ligand-receptor interactions on living cells and tissues. *Nature Biotechnol.* **30**, 997–1001 (2012).
77. Herzog, F. *et al.* Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry. *Science* **337**, 1348–1352 (2012).
- This study pioneered the use of chemical crosslinking to reveal the topology of an important phosphatase complex.**
78. Liu, F., Rijkers, D. T. S., Post, H. & Heck, A. J. R. Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nature Methods* **12**, 1179–1184 (2015).
79. Navare, A. T. *et al.* Probing the protein interaction network of *Pseudomonas aeruginosa* cells by chemical cross-linking mass spectrometry. *Structure* **23**, 762–773 (2015).
80. Makowski, M. M., Willems, E., Jansen, P. W. T. C. & Vermeulen, M. Cross-linking immunoprecipitation-MS (XIP-MS): topological analysis of chromatin-associated protein complexes using single affinity purification. *Mol. Cell. Proteomics* **15**, 854–865 (2016).
81. Shi, Y. *et al.* A strategy for dissecting the architectures of native macromolecular assemblies. *Nature Methods* **12**, 1135–1138 (2015).
82. Auferheide, A. *et al.* Structural characterization of the interaction of Ubp6 with the 26S proteasome. *Proc. Natl Acad. Sci. USA* **112**, 8626–8631 (2015).
83. Mahamid, J. *et al.* Visualizing the molecular sociology at the HeLa cell nuclear periphery. *Science* **351**, 969–972 (2016).
84. Engen, J. R. Analysis of protein conformation and dynamics by hydrogen/deuterium exchange MS. *Anal. Chem.* **81**, 7870–7875 (2009).
85. Wang, L. & Chance, M. R. Structural mass spectrometry of proteins using hydroxyl radical based protein footprinting. *Anal. Chem.* **83**, 7234–7241 (2011).
86. Savitskii, M. M. *et al.* Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science* **346**, 1255784 (2014).
- In this paper, isobaric chemical labelling was used to measure the proportion of proteins that bound to a drug as a function of temperature, on a proteome-wide scale.**
87. Feng, Y. *et al.* Global analysis of protein structural changes in complex proteomes. *Nature Biotechnol.* **32**, 1036–1044 (2014).
88. Pauling, L., Itano, H. A., Singer, S. J. & Wells, I. C. Sickle cell anemia, a molecular disease. *Science* **110**, 543–548 (1949).
89. Picotti, P. *et al.* A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature* **494**, 266–270 (2013).
90. Andreux, P. A. *et al.* Systems genetics of metabolism: the use of the BXD murine reference panel for multiscale integration of traits. *Cell* **150**, 1287–1299 (2012).
91. Wu, Y. *et al.* Multilayered genetic and omics dissection of mitochondrial activity in a mouse reference population. *Cell* **158**, 1415–1430 (2014).
92. Williams, E. G. *et al.* Systems proteomics of liver mitochondria function. *Science* **352**, aad0189 (2016).
- A demonstration of the combined use of proteomics and genetics to interrogate mitochondrial function.**
93. Mertins, P. *et al.* Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534**, 55–62 (2016).
- This analysis of breast cancer tissues revealed that proteomics is almost on a par with transcriptomics in terms of achievable depth of coverage of gene expression.**
94. Carr, S. A. *et al.* Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. *Mol. Cell. Proteomics* **13**, 907–917 (2014).
95. Rifai, N., Gillette, M. A. & Carr, S. A. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nature Biotechnol.* **24**, 971–983 (2006).
96. Geyer, P. E. *et al.* Plasma proteome profiling to assess human health and disease. *Cell Syst.* **2**, 185–195 (2016).
97. Surinova, S. *et al.* Prediction of colorectal cancer diagnosis based on circulating plasma proteins. *EMBO Mol. Med.* **7**, 1166–1178 (2015).
98. Liu, Y. *et al.* Quantitative variability of 342 plasma proteins in a human twin population. *Mol. Syst. Biol.* **11**, 786 (2015).
99. Bandura, D. R. *et al.* Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Anal. Chem.* **81**, 6813–6822 (2009).
100. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnol.* **26**, 1367–1372 (2008).
101. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods* **13**, 731–740 (2016).
102. MacLean, B. *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968 (2010).
103. Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11**, 0111.016717 (2012).
104. Röst, H. L. *et al.* OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nature Biotechnol.* **32**, 219–223 (2014).
105. Tsou, C.-C. *et al.* DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nature Methods* **12**, 258–264 (2015).
106. Olsen, J. V. *et al.* Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci. Signal.* **3**, ra3 (2010).
107. Smith, L. M., Kelleher, N. L. & The Consortium for Top Down Proteomics. Proteoform: a single term describing protein complexity. *Nature Methods* **10**, 186–187 (2013).
108. Olsen, J. V. *et al.* Higher-energy C-trap dissociation for peptide modification analysis. *Nature Methods* **4**, 709–712 (2007).
109. Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J. & Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl Acad. Sci. USA* **101**, 9528–9533 (2004).
110. Zubarev, R. A. & Makarov, A. Orbitrap mass spectrometry. *Anal. Chem.* **85**, 5288–5296 (2013).
111. Picotti, P. & Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nature Methods* **9**, 555–566 (2012).
112. Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S. & Coon, J. J. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol. Cell. Proteomics* **11**, 1475–1488 (2012).
113. Chapman, J. D., Goodlett, D. R. & Masselon, C. D. Multiplexed and data-independent tandem mass spectrometry for global proteome profiling. *Mass Spectrom. Rev.* **33**, 452–470 (2014).
114. Rosenberger, G. *et al.* A repository of assays to quantify 10,000 human proteins by SWATH-MS. *Sci. Data* **1**, 140031 (2014).
115. Meier, F. *et al.* Parallel accumulation–serial fragmentation (PASEF): multiplying sequencing speed and sensitivity by synchronized scans in a trapped ion mobility device. *J. Proteome Res.* **14**, 5378–5387 (2015).
116. Ow, S. Y. *et al.* iTRAQ underestimation in simple and complex mixtures: “the good, the bad and the ugly”. *J. Proteome Res.* **8**, 5347–5355 (2009).
117. Ting, L., Rad, R., Gygi, S. P. & Haas, W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nature Methods* **8**, 937–940 (2011).
118. Wühr, M. *et al.* The nuclear proteome of a vertebrate. *Curr. Biol.* **25**, 2663–2671 (2015).
119. Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–2526 (2014).
120. Ludwig, C., Claassen, M., Schmidt, A. & Aebersold, R. Estimation of absolute protein quantities of unlabeled samples by selected reaction monitoring mass spectrometry. *Mol. Cell. Proteomics* **11**, M111.013987 (2012).

Acknowledgements We thank M. Faini and R. Ciuffa for help in preparing the figures and Y. Liu for help in compiling the literature citations. M. Hein provided inspiration for this Review, read the manuscript critically and helped with preparing the figures, as did F. Hosp, P. Geyer and S. Beck. We thank members of our groups for critical discussions.

Author information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this paper at go.nature.com/2bqo2n8. Correspondence should be addressed to R.A. (aebersold@imsb.biol.ethz.ch) or M.M. (mmann@biochem.mpg.de).

Locus coeruleus and dopaminergic consolidation of everyday memory

Tomonori Takeuchi^{1*}, Adrian J. Duszakiewicz^{1*}, Alex Sonneborn^{2*}, Patrick A. Spooner¹, Miwako Yamasaki³, Masahiko Watanabe³, Caroline C. Smith², Guillén Fernández⁴, Karl Deisseroth⁵, Robert W. Greene^{2,6} & Richard G. M. Morris^{1,7}

The retention of episodic-like memory is enhanced, in humans and animals, when something novel happens shortly before or after encoding. Using an everyday memory task in mice, we sought the neurons mediating this dopamine-dependent novelty effect, previously thought to originate exclusively from the tyrosine-hydroxylase-expressing (TH⁺) neurons in the ventral tegmental area. Here we report that neuronal firing in the locus coeruleus is especially sensitive to environmental novelty, locus coeruleus TH⁺ neurons project more profusely than ventral tegmental area TH⁺ neurons to the hippocampus, optogenetic activation of locus coeruleus TH⁺ neurons mimics the novelty effect, and this novelty-associated memory enhancement is unaffected by ventral tegmental area inactivation. Surprisingly, two effects of locus coeruleus TH⁺ photoactivation are sensitive to hippocampal D₁/D₅ receptor blockade and resistant to adrenoceptor blockade: memory enhancement and long-lasting potentiation of synaptic transmission in CA1 *ex vivo*. Thus, locus coeruleus TH⁺ neurons can mediate post-encoding memory enhancement in a manner consistent with possible co-release of dopamine in the hippocampus.

Studies of memory for over a century¹ indicate that there is substantial forgetting within a day². Everyday memory includes many episodic-like memories that we may form automatically^{3,4}. Most are forgotten, others retained for longer such that they can then be subject to stabilization in the neocortex via systems consolidation^{5,6}. Initial retention occurs when something novel or categorically relevant happens shortly before or after the time of memory encoding, as in ‘flashbulb memory’^{7,8}. Prospective studies of novelty-associated enhancement of retention in animals point to possible mechanisms^{9–14}, one suggestion being that a novelty signal from the ventral tegmental area (VTA) to the hippocampus is causally important^{15,16}. Pharmacological studies of protein-synthesis-dependent long-term potentiation and depression (LTP/LTD) support the concept of dopamine-dependent enhancement of persistence^{13,17–19}, but do not identify the neuronal source of dopamine. We anticipated that TH⁺ neurons in the VTA (VTA-TH⁺) would be critical, but the possibility that locus coeruleus (LC)-TH⁺ neurons can sometimes release dopamine as well as noradrenaline^{20–22} led to our broadening the project to include LC-TH⁺ as well as VTA-TH⁺ neurons.

Novelty enhances memory persistence

We used 120 mice and began by checking that enhancement of memory retention by environmental novelty, previously shown in rats^{12,13}, could also be observed in mice. This involved our ‘everyday’ memory task—a model of ‘episodic-like’ or ‘one-shot’ memory²³—using a smaller arena for mice who searched for food reward in a sandwell whose location varied across days (Fig. 1a and Extended Data Fig. 1a). Male Th-Cre mice encoded the changing daily location of reward in ‘sample’ trials and learned to display effective memory in later daily ‘choice’ trials (five sandwells). The data, plotted as a five-alternative forced-choice performance index, revealed stable choice performance (circa 80%) across weeks of testing (Fig. 1a and Extended Data Fig. 1b, c).

A control test (sessions 61–65) established the absence of olfactory artefacts. Memory was tested using a counterbalanced series of occasional unrewarded probe tests that sometimes followed the sample trial(s) at different intervals. Reward availability on sample trials (memory encoding) was varied (low and high reward). Effective memory at 1 h (memory retrieval) was displayed in probe tests for even the smallest reward, but forgetting across 24 h showed an inverse relationship with reward magnitude (Fig. 1b and Extended Data Fig. 1d). A key finding was that unexpected novelty for 5 min—exploration of a box with an unfamiliar floor surface placed inside the event arena 30 min after the encoding trial—prolonged spatial memory to 24 h (Fig. 1b). This post-encoding novelty was environmental and not a change in reward expectancy.

Memory retention in this task is impaired by a drug targeting D₁/D₅ receptors in hippocampus in rats¹³. In our mice, the D₁/D₅ receptor antagonist SCH23390 (SCH) blocked novelty-induced enhancement of memory persistence, whereas the β -adrenoceptor antagonist propranolol (Prop) had no effect (Fig. 1c and Extended Data Fig. 1e).

TH⁺ neurons show a novelty response

The next step was to identify whether TH⁺ neurons are activated by environmental novelty. Th-Cre mice were given stereotaxic injections of a Cre-inducible adeno-associated virus (AAV) carrying channel-rhodopsin-2 (ChR2) fused to enhanced yellow fluorescent protein (eYFP) into the VTA or LC and implanted with an ‘optrode’ (Fig. 2a and Extended Data Fig. 2). Several weeks later, brief low-frequency blue light illumination was used to identify ChR2-eYFP-positive neurons in awake mice. TH⁺ neurons were identified (Fig. 2a and Extended Data Fig. 3a, b), and recording then continued without light pulses as the mice explored both familiar (regular visits) and novel environments (rare visits, each visit with a novel floor surface) for 5 min in a counterbalanced sequence (Fig. 2b). Representative raster plots

¹Centre for Cognitive and Neural Systems, Edinburgh Neuroscience, The University of Edinburgh, 1 George Square, Edinburgh EH8 9JZ, UK. ²University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75390, USA. ³Department of Anatomy, Hokkaido University Graduate School of Medicine, Sapporo, Hokkaido, 060-8638, Japan. ⁴Donders Institute for Brain, Cognition, and Behaviour, Radboud University Medical Centre, Nijmegen, 6525 EZ, The Netherlands. ⁵Departments of Psychiatry and Behavioral Sciences and of Bioengineering, Stanford University, Stanford, California 94305, USA. ⁶International Institute for Integrative Sleep Medicine (WPI-IIMS), University of Tsukuba, Tsukuba, 305-8575, Japan. ⁷Instituto de Neurociencias, CSIC-UMH, Alicante, 03550, Spain.

*These authors contributed equally to this work.

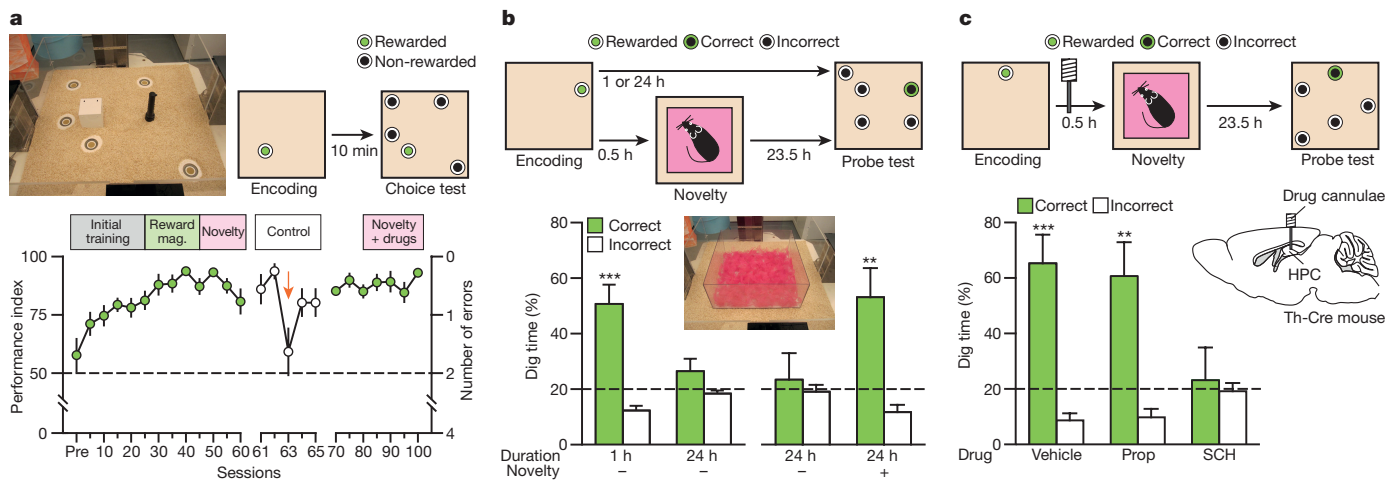


Figure 1 | Novelty exploration after memory encoding enhances memory retention. **a**, Everyday spatial memory task in event arena. Mice ($n = 13$; 100 sessions (Ss)) acquired stable performance (S26–S100: $F_{14,168} = 1.68$, $P > 0.05$). Non-encoding control session performance (S63, orange arrow) dropped to 59.6% (S61–S65: $F_{4,48} = 3.63$, $P < 0.05$; S63: t -test versus chance, $t_{12} < 1$). Green circles, 5S average; white circles, 1S average; Pre, pre-training. **b**, Memory at 1 h (probe test) declined to chance at 24 h (1 h versus 24 h: $t_{12} = 2.94$, $P < 0.05$; 1 h versus chance: $t_{12} = 4.44$, $P < 0.001$). Novelty 30 min after encoding resulted in memory at 24 h

show novelty affected light-responsive neurons in both VTA-TH⁺ and LC (Fig. 2c), with increased firing rates in most VTA-TH⁺ (12/15 neurons) and all LC-TH⁺ (10/10) neurons (Fig. 2d). Firing rate modulation was quantitatively larger for LC-TH⁺ neurons, above the usual tonic firing rate²⁴, and it habituated monotonically over time (Fig. 2e and Extended Data Fig. 3c; ref. 25). VTA-TH⁺ and LC-TH⁺ neurons showed more frequent bursts in novel environments with a within-burst spike frequency of 12.8–53.4 Hz (Extended Data Fig. 3d, e). This bursting pattern was, by design, used to determine an effective optogenetic stimulation protocol for both *in vivo* and *ex vivo* experiments.

LC projects extensively to hippocampus

Previous retrograde tracing and immunohistochemical work in the rat has shown that only a subset of VTA afferents to the hippocampus are dopaminergic²⁶. Our cell-type-specific anterograde tracing revealed a paucity of projections from VTA-TH⁺ neurons to dorsal hippocampus, but substantial projections from LC-TH⁺ neurons (Fig. 3a–f). This involved unilateral injection of the Cre-inducible eYFP virus into VTA or LC of Th-Cre mice. Most TH⁺ neurons in VTA and LC expressed eYFP (Extended Data Fig. 2g). We first established co-localization of TH in eYFP⁺ axons in the dorsal hippocampus (Fig. 3b, d, top panels), and then co-localization of eYFP, TH and the noradrenergic transporter (NET) for LC axons but not VTA axons (Fig. 3b, d, bottom panels, and Extended Data Fig. 4). We quantified eYFP⁺ axons projecting to hippocampus, calculating both the area occupied by all eYFP-TH double-positive axons and the ratio of double-positive axons relative to all TH⁺ axons (Fig. 3e, f). An overwhelming proportion of axons came from LC-TH⁺ neurons with very few from VTA-TH⁺ neurons (see also ref. 21). Retrograde tracing with fluorescent retrobeads from dorsal hippocampus confirmed minimal transport to VTA, but double-labelled neurons in LC (retrobeads and TH-positivity; Extended Data Fig. 5).

LC activation mimics the novelty effect

The overarching aim of this project was to identify neuromodulatory neurons whose post-encoding firing promotes the consolidation of hippocampal-dependent memory. The data so far pointed to the need to change our focus to LC-TH⁺ neurons as likely mediators of environmental novelty, retaining a check on the impact of activating

(no novelty versus novelty: $t_{12} = 2.24$, $P < 0.05$; novelty: $t_{12} = 3.17$, $P < 0.01$). **c**, Blockade of hippocampal D₁/D₅ receptor (SCH) but not β -adrenoceptors (Prop) during novelty abolished 24 h memory ($F_{2,24} = 3.83$, $P < 0.05$; vehicle: $t_{12} = 4.38$, $P < 0.001$; Prop: $t_{12} = 3.33$, $P < 0.01$). HPC, hippocampus; Prop, propranolol; SCH, SCH23390. *** $P < 0.01$, **** $P < 0.001$ versus chance. Dashed lines, chance. Means \pm s.e.m. The mouse brain in this figure has been reproduced with permission from Franklin, K. B. J. & Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates* 3rd ed, 691 (Academic, 2007).

VTA-TH⁺ neurons. Th-Cre mice were stereotactically prepared with bilateral injections of the ChR2-eYFP (ChR2⁺) or a no opsin control (eYFP) virus (ChR2⁻), accompanied by implantation of bilateral optic cannulae into both LC and VTA and bilateral drug cannulae targeting dorsal hippocampus (Fig. 4a). An optimum optogenetic burst stimulation frequency of 25 Hz was chosen (Fig. 4b) on the basis of our within-burst firing data, and both LC-TH⁺ and VTA-TH⁺ neurons could follow this frequency in awake mice (Extended Data Figs 3e and 6a). After training to a performance index of circa 75%, these mice showed effective memory in a 10-min probe test (Extended Data Fig. 6b, c). The stage was then set for examining whether 5 min optogenetic burst activation, scheduled 30 min after memory encoding, could mimic the beneficial effects of environmental novelty on memory retention at 24 h (Fig. 4c).

The key finding was the striking persistence of memory over 24 h when 5 min of post-encoding (30 min) intermittent burst stimulation of LC-TH⁺ neurons with blue light was given to the ChR2⁺ mice in their home cages. Tested 24 h after weak memory encoding, LC-activated ChR2⁺ animals remembered the location of the sample sandwell sampled 30 min before light activation (Fig. 4d left; ChR2⁺ in LC-on). Memory for 24 h was not observed without LC light stimulation (ChR2⁺ in off), nor in separate ChR2⁻ controls. Light-activation of VTA, in the same ChR2⁺ animals, induced only a non-significant trend favouring some memory at 24 h, but we observed a similar trend in ChR2⁻ mice (Fig. 4d right, compare ChR2⁺ and ChR2⁻ in VTA-on). Not only did these upward trends not differ from chance, they also did not differ from each other, suggesting that the trend is unrelated to light activation of ChR2-positive neurons in VTA (for example, light-induced temperature changes²⁷).

We were therefore confronted by the paradox that light activation of LC enhances retention (Fig. 4d) but intrahippocampal infusion of a D₁/D₅ receptor antagonist during behavioural novelty blocks it (Fig. 1c). Accordingly, we examined the impact of post-encoding microinfusion of catecholaminergic antagonists into the hippocampus during optogenetic activation of LC (Fig. 4e). We confirmed that light activation of LC in ChR2⁺ mice enhanced 24-h spatial memory (Fig. 4f; ChR2⁺ in LC-on with vehicle) but this enhancement was blocked by intrahippocampal infusion of SCH (ChR2⁺ in LC-on with SCH) but not by Prop (ChR2⁺ in LC-on with Prop; see also Extended Data

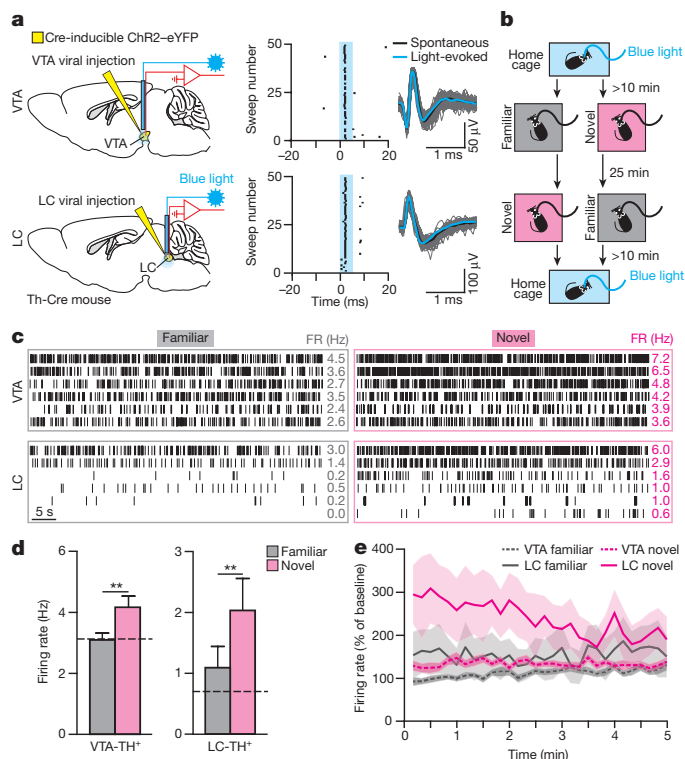


Figure 2 | LC-TH⁺ neurons show stronger modulation by novelty than VTA-TH⁺ neurons. **a**, Viral injection and optrode implantation. Putative VTA-TH⁺ and LC-TH⁺ neurons responded to blue light (blue). **b**, Behavioural protocol. **c**, Raster plot of VTA-TH⁺ (top) and LC-TH⁺ neurons (bottom) in familiar (left) and novel (right) environments. FR, firing rate. **d**, FRs of VTA-TH⁺ ($n = 15$ neurons, 5 mice) and LC-TH⁺ ($n = 10$ neurons, 3 mice) neurons were higher in the novel environment (VTA-TH⁺: $t_{14} = 4.30$, $P < 0.01$; LC-TH⁺: $t_9 = 3.46$, $P < 0.01$). Dashed lines, baseline. **e**, LC-TH⁺ neurons showed stronger modulation by novelty than VTA-TH⁺ neurons (brain area \times condition interaction, $F_{1,23} = 15.20$, $P < 0.001$). LC-TH⁺ but not VTA-TH⁺ neurons displayed habituation to novelty (LC novel, $F_{29,261} = 1.70$, $P < 0.05$; VTA novel, $F_{29,406} = 1.23$, $P > 0.05$). ** $P < 0.01$, paired t -test. Means \pm s.e.m. The mouse brain in this figure has been reproduced with permission from Franklin, K. B. J. & Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates* 3rd ed, 691 (Academic, 2007).

Fig. 6d). This raises the possibility of LC-TH⁺ terminals co-releasing dopamine in hippocampus, or of heterodimerization between noradrenaline and dopamine receptors.

LC activation enhances synaptic efficacy

To explore one possible mechanism of this enhancement of memory retention, *ex vivo* electrophysiological experiments examined the response of CA1 pyramidal neurons to CA3 Schaffer collateral synaptic input (Fig. 5a). Three weeks before obtaining the slices, bilateral injections of a Cre-inducible ChR2-eYFP virus were made into LC of Th-Cre mice (Extended Data Fig. 7a), in keeping with the recently described ‘output-defined elements’ concept²⁸.

Following three trains of burst optogenetic stimulation of hippocampal LC-TH⁺ axons (Extended Data Fig. 7b), the CA3–CA1 excitatory postsynaptic currents (EPSCs) gradually increased by 55% over ~ 30 min (Fig. 5b, light on), an increase unaffected by the presence of the α - and β -adrenoceptors antagonists prazosin (Praz) and Prop. In contrast, there was no increase following optogenetic LC-TH⁺ activation in the presence of SCH, revealing a pattern of EPSC potentiation consistent with mediation by a dopaminergic mechanism.

LTP at CA3–CA1 synapses was then examined using theta-burst stimulation to induce LTP. Hippocampal LC-TH⁺ axons were

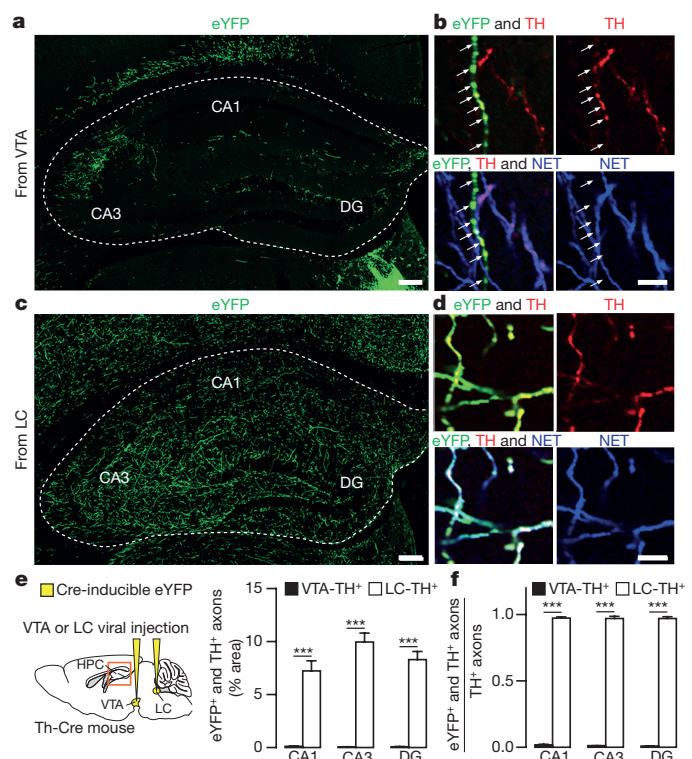


Figure 3 | TH⁺ axons in the hippocampus originate from LC-TH⁺ neurons. **a–d**, Representative coronal sections show overall distribution of eYFP⁺ axons from VTA (**a**) and LC (**c**) in dorsal hippocampus. Triple immunofluorescence for eYFP (green), TH (red) and NET (blue) shows co-labelling of eYFP⁺ VTA axons with TH (**b**, top) but not with NET (**b**, bottom; arrows), co-labelling of eYFP⁺ LC axons with TH (**d**, top) and NET (**d**, bottom). **e**, **f**, Quantification of area occupied by eYFP and TH double-positive axons relative to all TH⁺ axons (f) in CA1, CA3 and DG ($n = 9$ slices, 3 mice per group). Both measures indicate stronger TH⁺ projections from LC than from VTA in CA1 (area: $t_{16} = 7.4$, $P < 0.001$; ratio: $t_{16} = 104.1$, $P < 0.001$), CA3 (area: $t_{16} = 11.7$, $P < 0.001$; ratio: $t_{16} = 59.0$, $P < 0.001$) and DG (area: $t_{16} = 10.8$, $P < 0.001$; ratio: $t_{16} = 76.4$, $P < 0.001$). *** $P < 0.001$, paired t -test. Means \pm s.e.m. The mouse brain in this figure has been reproduced with permission from Franklin, K. B. J. & Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates* 3rd ed, 691 (Academic, 2007).

selectively activated with a protocol closely mimicking LC-TH⁺ firing patterns recorded during novelty exploration (Extended Data Fig. 7c). LTP differed in magnitude across four conditions (Fig. 5c). Theta-burst stimulation alone induced LTP by 29% at 45 min (light off + LTP) relative to a no-LTP baseline; by 59% when combined with optogenetic LC-TH⁺ activation (light on + LTP); but by was blocked by SCH (light on + LTP with SCH). Taken together, these findings indicate that depolarization of hippocampal LC-TH⁺ axons by optogenetic stimulation can enhance synaptic transmission, and that a physiologically realistic pattern potentiates LTP at CA3–CA1 synapses in a manner consistent with release of dopamine from hippocampal LC-TH⁺ terminals.

VTA blockade has no impact on novelty effect

Last, we attempted to block LC-TH⁺ neurons during novelty to see if memory enhancement disappeared. Acute electrophysiological studies showed that firing of LC-TH⁺ neurons expressing archaerhodopsin (eArch3.0)–eYFP could be successfully inhibited by light over 5 min (Extended Data Fig. 8). Unfortunately, TH⁺ neurons always displayed a substantial ‘rebound’ of firing when the light was turned off²⁹. Were this rebound to occur in a behavioural study of memory, the rebound would probably overcome the earlier 5 min period of neuronal quiescence.

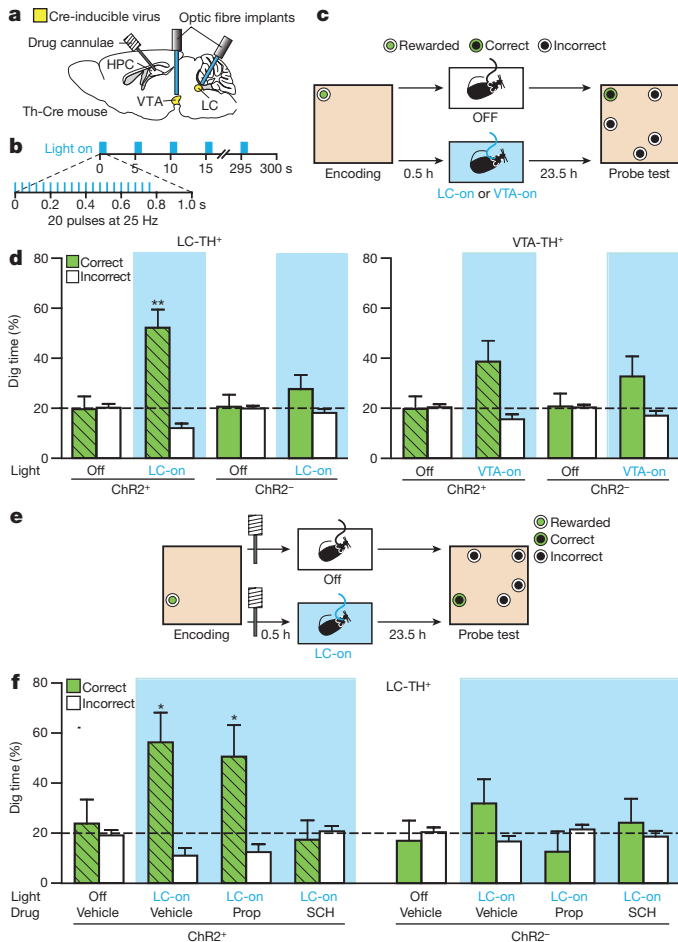


Figure 4 | Optogenetic activation of LC-TH⁺ neurons enhances memory persistence. **a**, Viral injection, optic fibre and drug cannulae implantations. Th-Cre mice injected with a Cre-inducible ChR2-eYFP AAV (ChR2⁺, $n = 8$) or a control eYFP AAV (ChR2⁻, $n = 6$) into LC and VTA. **b**, Optogenetic burst protocol used in the event arena. **c**, Design for optogenetic mimicry experiment. **d**, Left, LC-TH⁺ neuron photostimulation (LC-on) 30 min after encoding enhanced 24 h memory in ChR2⁺ animals but not in ChR2⁻ controls (group × condition interaction, $F_{1,12} = 5.66$, $P < 0.05$; ChR2⁺ in LC-on versus chance: $t_7 = 4.38$, $P < 0.01$). Right, VTA-TH⁺ neuron photostimulation (VTA-on) caused a trend for enhanced memory that did not differ between groups (group × condition interaction, $F_{1,12} = 0.33$, $P = 0.58$; ChR2⁺ in VTA-on: $t_7 = 2.22$, $P = 0.062$; ChR2⁻ in VTA-on: $t_5 = 1.56$, $P = 0.18$). **e**, Design for optogenetic LC activation experiment with pharmacological interventions. **f**, Blockade of hippocampal D₁/D₅ receptor (SCH) but not β -adrenoceptors (Prop) during LC-TH⁺ neuron photostimulation abolished the effect of LC photostimulation on memory persistence in ChR2⁺ mice (group effect, $F_{1,12} = 5.01$, $P < 0.05$; condition effect in ChR2⁺, $F_{3,21} = 3.18$, $P < 0.05$; in ChR2⁻: $F_{3,15} < 1$). ChR2⁺ mice showed good memory with post-encoding LC-on in presence of vehicle or Prop, but not in presence of SCH, or without light stimulation (orthogonal comparison: $F_{1,21} = 9.23$, $P < 0.01$; LC-on with vehicle in ChR2⁺: $t_7 = 3.07$, $P < 0.05$; LC-on with Prop in ChR2⁺: $t_7 = 2.41$, $P < 0.05$). Prop, propranolol; SCH, SCH23390. * $P < 0.05$, ** $P < 0.01$ versus chance. Dashed lines, chance. Means \pm s.e.m. The mouse brain in this figure has been reproduced with permission from Franklin, K. B. J. & Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates* 3rd ed, 691 (Academic, 2007).

At this point, we had Th-Cre mice expressing eArch3.0-eYFP or eYFP in LC trained with comparable levels of task performance to those of the earlier experiments (Extended Data Fig. 9a, b). While we were unable to use eArch3.0, the mice had bilateral drug cannulae targeting VTA. In separate acute electrophysiology experiments, the sodium channel blocker lidocaine successfully inhibited VTA

neuronal firing (Fig. 6a). We therefore examined the impact of intra-VTA infusion of lidocaine during novelty exploration. As expected, novelty enhancement was maintained (Fig. 6b and Extended Data Fig. 9c). The novelty effect was, however, sensitive to the α_2 -adrenoceptor agonist clonidine (Fig. 6c and Extended Data Fig. 9c) that inhibits LC neurons³⁰ (O. Eschenko, personal communication, 2016). The mice were given a series of counterbalanced probe tests. One test showed normal forgetting over 24 h with low-reward encoding (No novelty with vehicle), another the usual enhanced retention induced by environmental novelty (novelty with vehicle). Critically, systemic clonidine blocked the effect of novelty on 24 h memory retention (novelty with clonidine).

Discussion

This study sought to identify neuromodulatory neurons that, in response to environmental novelty, trigger intracellular signal-transduction cascades within hippocampal neurons responsible for the initial consolidation of everyday memory^{15,31,32}. The ‘hippocampal-VTA loop’ model of dopaminergic consolidation¹⁵ postulates that novelty-associated enhancement of hippocampal-dependent memory is mediated by a subiculum–accumbens–pallidum–VTA–CA1 pathway, an idea supported by both human and animal studies^{16,33}. Our data indicate, however, that VTA-TH⁺ neurons display weak anatomical connectivity with hippocampus, are only slightly activated by environmental novelty, and their pharmacological blockade during novelty has no effect on memory enhancement. In contrast, LC-TH⁺ neurons are more responsive to environmental novelty, show habituation of enhanced firing over 5 min, have extensive connectivity with hippocampus, and display post-encoding optogenetic activation that successfully enhances memory retention. Clonidine, which decreases neuronal activity in LC³⁰ but not in VTA³⁴, prevents this novelty effect. At the synaptic level *ex vivo*, selective optogenetic activation of LC-TH⁺ axons in hippocampus mediates a D₁/D₅ receptor-sensitive enhancement of synaptic transmission and electrically evoked LTP at CA3–CA1 synapses, consistent with the idea that LC-TH⁺ neurons might co-release dopamine^{20,21}. Transmitter co-release of glutamate and GABA (γ -aminobutyric acid) from TH⁺ neurons has been reported (substantia nigra (SN)/VTA-TH⁺ terminals^{35,36}). The ‘hippocampal-VTA loop’ model has been modified to recognize this possibility³⁷, rescuing dopamine receptor dependence at the expense of recognizing a different circuit for mediating the impact of environmental novelty on hippocampal memory.

Importantly, environmental novelty differs from changes in reward expectancy or magnitude. Reward expectancy is a critical component of the execution of learned actions until they become habitual³⁸. Longstanding data point to the role of SN/VTA system in processing unexpected reward in striatal^{39–41} and hippocampal tasks⁴². Activation of VTA-TH⁺ axons can bidirectionally modulate CA3–CA1 synaptic responses *ex vivo*⁴³, but in our *in vivo* study, post-encoding VTA-TH⁺ activation was without a behavioural effect. Environmental novelty is when something unexpected happens unrelated to reward during ongoing behaviour. It is likely to affect a distinct neuromodulatory system, but one with extensive connectivity to the hippocampus where it could activate memory processes including the synthesis and capture of plasticity-related proteins³¹. In this way, unexpected environmental novelty could enhance memory retention. The noradrenergic system of the LC, with its diverse projections to numerous brain areas⁴⁴, has long been implicated in novelty, arousal and cognition^{24,45,46}, and its firing is tied to distinct up and down states during sleep⁴⁷. It appears that the time over which LC neurons exert their effects is much less precise than for VTA, in keeping with the synaptic tagging-and-capture hypothesis^{48–50}. This hypothesis has functional implications that extend well beyond the domain of LTP, affecting the retention of events via neural activity happening before and after the precise moment that encoding occurs

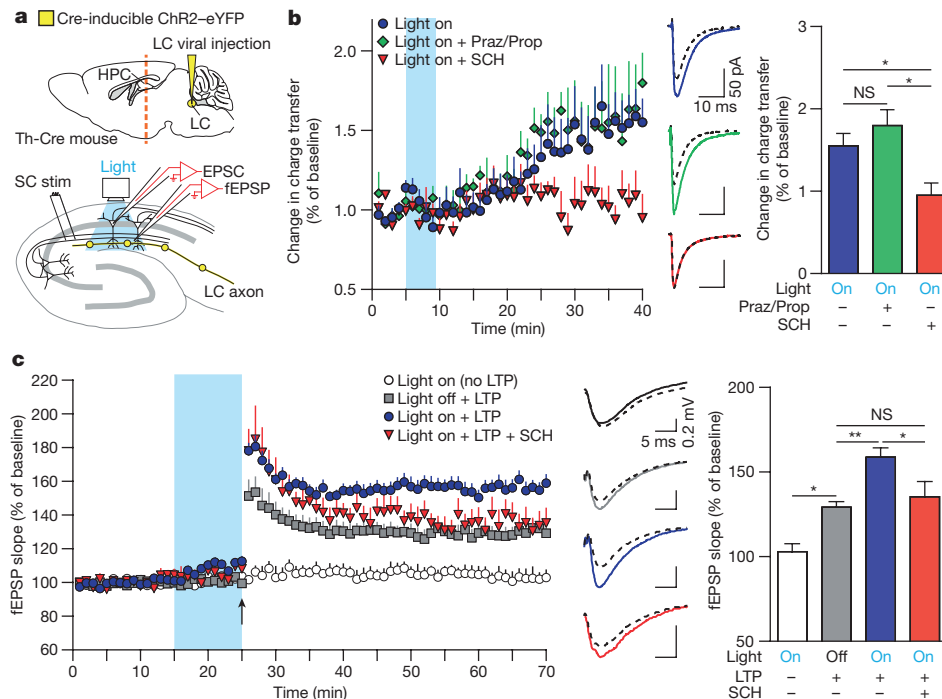


Figure 5 | Optogenetic activation of LC-TH⁺ axons enhances hippocampal synaptic function. **a**, Hippocampal slice physiology. Orange line, slice plane. **b**, Left, potentiation of Schaffer collateral (SC)-evoked EPSCs from CA1 pyramidal neurons after strong light activation (blue) of hippocampal LC-TH⁺ axons (light on, $n = 5$) was unaffected by adrenoceptor antagonists (light on with Praz/Prop, $n = 4$) but blocked by D₁/D₅ receptor antagonist (light on with SCH, $n = 5$) (conditions \times time interaction, $F_{7.8, 42.8} = 2.50$, $P < 0.05$, Greenhouse–Geisser correction). Middle, exemplar EPSCs from CA1 pyramidal neurons. Dashed lines, baseline EPSCs; Continuous, EPSCs 30–35 min after light onset. Right, mean EPSCs 35 min after light onset, showing effect of SCH but no effect of Praz/Prop ($F_{2,11} = 7.20$, $P < 0.05$). **c**, Left, fEPSP responses to weak theta-burst stimulation (arrow) with or without optogenetic activation of hippocampal LC-TH⁺ axons (blue). No synaptic potentiation without

theta-burst (light on (no LTP), $n = 6$), but with it, an increase in synaptic strength lasting >45 min (light off + LTP, $n = 6$) that was significantly enhanced by a weak physiologically relevant optogenetic stimulation of LC-TH⁺ axons (light on + LTP, $n = 11$). SCH blocked enhancement of LTP (light on + LTP with SCH, $n = 5$) (conditions \times time interaction, $F_{19.5/155.8} = 3.01$, $P < 0.001$). Middle, fEPSPs: baseline (dashed lines) and 40–45 min after theta-burst stimulation (continuous lines). Right, mean fEPSP slopes 45 min after theta-burst stimulation, shows blockade of optogenetic augmentation of LTP by SCH ($F_{3,24} = 17.00$, $P < 0.001$). SCH, SCH23390 or SCH39166 (see Methods). NS, not significant. * $P < 0.05$, ** $P < 0.01$, Tukey's HSD test. Means \pm s.e.m. The mouse brain in this figure has been reproduced with permission from Franklin, K. B. J. & Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates* 3rd ed, 691 (Academic, 2007).

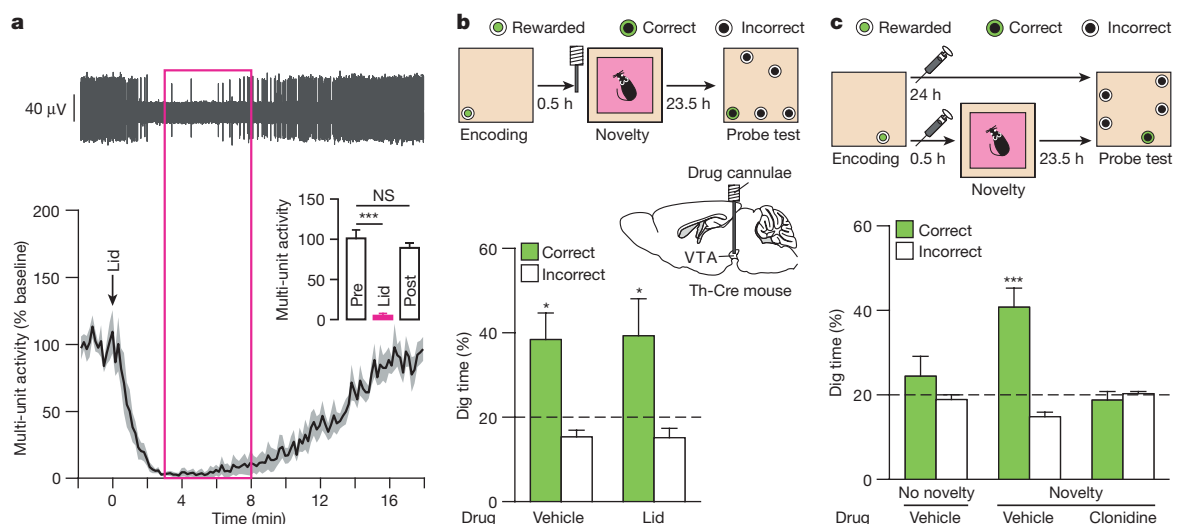


Figure 6 | Pharmacological inhibition of VTA has no impact on the novelty effect. **a**, Microinfusion of lidocaine (Lid) into VTA blocks multi-unit activity (example trace and population data; $n = 8$ traces, 4 mice; Pre versus Lid: $t_7 = 8.42$, $P < 0.001$; Pre versus Post: $t_7 = 1.42$, $P > 0.05$). Magenta box, novelty period. **b**, Lidocaine into the VTA before novelty had no effect on memory enhancement ($n = 15$ mice) (vehicle versus Lid: $t_{14} < 1$; vehicle versus chance: $t_{14} = 2.95$, $P < 0.05$; Lid: $t_{14} = 2.19$, $P < 0.05$).

c, Systemic injection of α_2 -adrenoceptor agonist clonidine before novelty abolished novelty effect ($F_{2,28} = 7.70$, $P < 0.01$; novelty with vehicle: $t_{14} = 4.62$, $P < 0.001$). NS, not significant. * $P < 0.05$, *** $P < 0.001$. Dashed lines, chance. Means \pm s.e.m. The mouse brain in this figure has been reproduced with permission from Franklin, K. B. J. & Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates* 3rd ed, 691 (Academic, 2007).

(as in flashbulb memory⁷). In this way, the retention of everyday experience is modulated over time and not just at the precise moment that individual events are encoded.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 September 2015; accepted 24 July 2016.

Published online 7 September 2016.

- Ebbinghaus, H. *Memory: A Contribution to Experimental Psychology* (Dover, 1885/1964).
- Wixted, J. T. The psychology and neuroscience of forgetting. *Annu. Rev. Psychol.* **55**, 235–269 (2004).
- Marr, D. Simple memory: a theory for archicortex. *Phil. Trans. R. Soc. Lond. B* **262**, 23–81 (1971).
- Morris, R. G. M. Elements of a neurobiological theory of hippocampal function: the role of synaptic plasticity, synaptic tagging and schemas. *Eur. J. Neurosci.* **23**, 2829–2846 (2006).
- Squire, L. R. Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychol. Rev.* **99**, 195–231 (1992).
- Dudai, Y. & Morris, R. G. M. in *Brain, Perception and Memory: Advances in Cognitive Neuroscience* (ed. Bolhuis, J. J.) 147–162 (Oxford Univ. Press, 2000).
- Brown, R. & Kulik, J. Flashbulb memories. *Cognition* **5**, 73–99 (1977).
- Dunsmoor, J. E., Murty, V. P., Davachi, L. & Phelps, E. A. Emotional learning selectively and retroactively strengthens memories for related events. *Nature* **520**, 345–348 (2015).
- Guzowski, J. F., McNaughton, B. L., Barnes, C. A. & Worley, P. F. Environment-specific expression of the immediate-early gene *Arc* in hippocampal neuronal ensembles. *Nature Neurosci.* **2**, 1120–1124 (1999).
- Li, S., Cullen, W. K., Anwyl, R. & Rowan, M. J. Dopamine-dependent facilitation of LTP induction in hippocampal CA1 by exposure to spatial novelty. *Nature Neurosci.* **6**, 526–531 (2003).
- Lemon, N. & Manahan-Vaughan, D. Dopamine D₁/D₅ receptors gate the acquisition of novel information through hippocampal long-term potentiation and long-term depression. *J. Neurosci.* **26**, 7723–7729 (2006).
- Moncada, D. & Viola, H. Induction of long-term memory by exposure to novelty requires protein synthesis: evidence for a behavioral tagging. *J. Neurosci.* **27**, 7476–7481 (2007).
- Wang, S. H., Redondo, R. L. & Morris, R. G. M. Relevance of synaptic tagging and capture to the persistence of long-term potentiation and everyday spatial memory. *Proc. Natl Acad. Sci. USA* **107**, 19537–19542 (2010).
- Takeuchi, T., Duszkiewicz, A. J. & Morris, R. G. M. The synaptic plasticity and memory hypothesis: encoding, storage and persistence. *Phil. Trans. R. Soc. B* **369**, 20130288 (2013).
- Lisman, J. E. & Grace, A. A. The hippocampal-VTA loop: controlling the entry of information into long-term memory. *Neuron* **46**, 703–713 (2005).
- Bunzeck, N., Guitart-Masip, M., Dolan, R. J. & Duzel, E. Pharmacological dissociation of novelty responses in the human brain. *Cereb. Cortex* **24**, 1351–1360 (2014).
- Frey, U., Matthies, H., Reymann, K. G. & Matthies, H. The effect of dopaminergic D1 receptor blockade during tetanization on the expression of long-term potentiation in the rat CA1 region *in vitro*. *Neurosci. Lett.* **129**, 111–114 (1991).
- Huang, Y. Y. & Kandel, E. R. D. D1/D5 receptor agonists induce a protein synthesis-dependent late potentiation in the CA1 region of the hippocampus. *Proc. Natl Acad. Sci. USA* **92**, 2446–2450 (1995).
- Lemon, N. & Manahan-Vaughan, D. Dopamine D₁/D₅ receptors contribute to de novo hippocampal LTD mediated by novel spatial exploration or locus coeruleus activity. *Cereb. Cortex* **22**, 2131–2138 (2012).
- Devoto, P. & Flore, G. On the origin of cortical dopamine: is it a co-transmitter in noradrenergic neurons? *Curr. Neuropharmacol.* **4**, 115–125 (2006).
- Smith, C. & Greene, R. W. CNS dopamine transmission mediated by noradrenergic innervation. *J. Neurosci.* **32**, 6072–6080 (2012).
- Hansen, N. & Manahan-Vaughan, D. Dopamine D₁/D₅ receptors mediate informational saliency that promotes persistent hippocampal long-term plasticity. *Cereb. Cortex* **24**, 845–858 (2014).
- Nakazawa, K. *et al.* Hippocampal CA3 NMDA receptors are crucial for memory acquisition of one-time experience. *Neuron* **38**, 305–315 (2003).
- Aston-Jones, G. & Bloom, F. E. Activity of norepinephrine-containing locus coeruleus neurons in behaving rats anticipates fluctuations in the sleep-waking cycle. *J. Neurosci.* **1**, 876–886 (1981).
- Vankov, A., Hervé-Minvielle, A. & Sara, S. J. Response to novelty and its rapid habituation in locus coeruleus neurons of the freely exploring rat. *Eur. J. Neurosci.* **7**, 1180–1187 (1995).
- Gasbarri, A., Verney, C., Innocenzi, R., Campana, E. & Pacitti, C. Mesolimbic dopaminergic neurons innervating the hippocampal formation in the rat: a combined retrograde tracing and immunohistochemical study. *Brain Res.* **668**, 71–79 (1994).
- Stujenske, J. M., Spellman, T. & Gordon, J. A. Modeling the spatiotemporal dynamics of light and heat propagation for *in vivo* optogenetics. *Cell Reports* **12**, 525–534 (2015).
- Lerner, T. N., Ye, L. & Deisseroth, K. Communication in neural circuits: tools, opportunities, and challenges. *Cell* **164**, 1136–1150 (2016).
- Mahn, M., Prigge, M., Ron, S., Levy, R. & Yizhar, O. Biophysical constraints of optogenetic inhibition at presynaptic terminals. *Nature Neurosci.* **19**, 554–556 (2016).
- Wu, M. F. *et al.* Locus coeruleus neurons: cessation of activity during catalepsy. *Neuroscience* **91**, 1389–1399 (1999).
- Frey, U. & Morris, R. G. M. Synaptic tagging: implications for late maintenance of hippocampal long-term potentiation. *Trends Neurosci.* **21**, 181–188 (1998).
- Kandel, E. R. The molecular biology of memory: cAMP, PKA, CRE, CREB-1, CREB-2, and CPEB. *Mol. Brain* **5**, 14 (2012).
- McNamara, C. G., Tejero-Cantero, A., Trouche, S., Campo-Urriza, N. & Dupret, D. Dopaminergic neurons promote hippocampal reactivation and spatial memory persistence. *Nature Neurosci.* **17**, 1658–1660 (2014).
- Grenhoff, J. & Svensson, T. H. Clonidine modulates dopamine cell firing in rat ventral tegmental area. *Eur. J. Pharmacol.* **165**, 11–18 (1989).
- Chuhma, N. *et al.* Dopamine neurons mediate a fast excitatory signal via their glutamatergic synapses. *J. Neurosci.* **24**, 972–981 (2004).
- Tritsch, N. X., Ding, J. B. & Sabatini, B. L. Dopaminergic neurons inhibit striatal output through non-canonical release of GABA. *Nature* **490**, 262–266 (2012).
- Otmakhova, N. A., Duzel, E., Deutch, A. Y. & Lisman, J. in *Intrinsically Motivated Learning in Natural and Artificial Systems* (eds Baldassarre, G. & Mirolli, M.) 235–254 (Springer, 2013).
- Everitt, B. J., Dickinson, A. & Robbins, T. W. The neuropsychological basis of addictive behaviour. *Brain Res. Brain Res. Rev.* **36**, 129–138 (2001).
- Schultz, W., Apicella, P., Ljungberg, T., Romo, R. & Scarnati, E. Reward-related activity in the monkey striatum and substantia nigra. *Prog. Brain Res.* **99**, 227–235 (1993).
- Tsai, H. C. *et al.* Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. *Science* **324**, 1080–1084 (2009).
- Steinberg, E. E. *et al.* A causal link between prediction errors, dopamine neurons and learning. *Nature Neurosci.* **16**, 966–973 (2013).
- Gruber, M. J., Ritchey, M., Wang, S. F., Doss, M. K. & Ranganath, C. Post-learning hippocampal dynamics promote preferential retention of rewarding events. *Neuron* **89**, 1110–1120 (2016).
- Rosen, Z. B., Cheung, S. & Siegelbaum, S. A. Midbrain dopamine neurons bidirectionally regulate CA3–CA1 synaptic drive. *Nature Neurosci.* **18**, 1763–1771 (2015).
- Schwarz, L. A. *et al.* Viral-genetic tracing of the input-output organization of a central noradrenergic circuit. *Nature* **524**, 88–92 (2015).
- Harley, C. Noradrenergic and locus coeruleus modulation of the perforant path-evoked potential in rat dentate gyrus supports a role for the locus coeruleus in attentional and memorial processes. *Prog. Brain Res.* **88**, 307–321 (1991).
- Sara, S. J. The locus coeruleus and noradrenergic modulation of cognition. *Nature Rev. Neurosci.* **10**, 211–223 (2009).
- Eschenko, O., Magri, C., Panzeri, S. & Sara, S. J. Noradrenergic neurons of the locus coeruleus are phase locked to cortical up-down states during sleep. *Cereb. Cortex* **22**, 426–435 (2012).
- Frey, U. & Morris, R. G. M. Synaptic tagging and long-term potentiation. *Nature* **385**, 533–536 (1997).
- Redondo, R. L. & Morris, R. G. M. Making memories last: the synaptic tagging and capture hypothesis. *Nature Rev. Neurosci.* **12**, 17–30 (2011).
- Rogerson, T. *et al.* Synaptic tagging during memory allocation. *Nature Rev. Neurosci.* **15**, 157–169 (2014).

Acknowledgements We thank M. Evans for Th-Cre mice; H. Rowe and D. Tse for pilot studies; M. Lo, H.-C. Tsai and C. Ramakrishnan for optogenetics training; UNC Viral Cores for producing AAVs; J. Cala, R. Fitzpatrick, J. Tulloch and T. Thai for technical support; R. Watson for animal care; N. Uchida and J. Cohen for assistance with optogenetic identification of TH⁺ neurons; P. Pedarzani, S. Canals, L. Genzel, M. Kroes, J. Lisman, D. Manahan-Vaughan, M. Munoz, T. Spiess-Jones, S.-H. Wang, O. Eschenko and E. Wood for scientific discussion. The study was funded by the European Research Council (G.F., R.G.M.M.: ERC-2010-AdG-268800-NEUROSCHEMA), UK Medical Research Council (A.J.D., R.G.M.M.), the European Commission's 7th Framework 2011 ICT Programme for Future Emerging Technologies (A.J.D., R.G.M.M.: 600725-GRIDMAP), Department Veterans Affairs, National Institutes of Health grant 5R01MH080297 (R.W.G.), and NIDA-T32-DA7290 Basic Science Training Program in Drug Abuse (principal investigator A. Eisch (A.S.)). The Instituto de Neurociencias at Alicante is 'Centre of Excellence Severo Ochoa'.

Author Contributions T.T. and A.J.D. designed with R.G.M.M. and conducted and analysed the behavioural, anatomical and *in vivo* electrophysiological experiments. T.T. and M.Y. designed with M.W. and conducted the tract-tracing; M.Y. analysed the data. A.S. and C.C.S. designed with R.W.G. and conducted the *ex vivo* electrophysiology. P.A.S. and A.J.D. constructed behavioural apparatus and optogenetic stimulation equipment. R.W.G., G.F. and R.G.M.M. secured the funding. K.D. offered advice and training to T.T. The manuscript was written by T.T., A.J.D., R.W.G. and R.G.M.M. All authors discussed the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.G.M.M. (R.G.M.Morris@ed.ac.uk) or R.W.G. (robertw.greene@utsouthwestern.edu).

Reviewer Information Nature thanks S. Thompson and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Animals. The subjects were Th-Cre knock-in heterozygous male mice backcrossed more than 20 times to the C57BL/6 strain (Th^{tm1(cre)Tc}; EM:00254)⁵¹ (behavioural and anatomical studies, $n = 71$ mice), Th-Cre transgenic heterozygous male mice on a mixed C57BL/6 and CD1 background (Tg(Th-cre)1Tmd)⁵² (for *ex vivo* hippocampal electrophysiology, $n = 42$), and C57BL/6 male mice (Charles River; $n = 7$). They were >8 weeks old at the start of the experiments, several of which continued for many months, and assigned to groups randomly. All mice were given water *ad libitum*, kept under a 12 h light/dark cycle (lights on 7:00); given food *ad libitum* in unit recording studies but food-restricted for event arena training (85% of free-feeding weight monitored daily throughout the study, after behavioural training). Behavioural testing was performed during the light phase of the cycle, and all critical tests were conducted 'blind'. All procedures were overseen by the University of Edinburgh Ethical Review Committee, compliant with the UK Animals (Scientific Procedures) Act 1986 and with the European Communities Council Directive of 24 November 1986 (86/609/EEC) legislation governing the maintenance of laboratory animals and their use in scientific experiments; and with guidelines of the Animal Welfare Committee of Hokkaido University; were approved by the animal care and use committee (IACUC) at the University of Texas Southwestern Medical Center and comply with federal regulations set forth by the National Institutes of Health.

Viral vectors. The Cre-inducible AAV were obtained from the University of North Carolina (UNC) Vector Core Facilities. The viral concentration was 8.0×10^{12} particles ml^{-1} for AAV5/EF1a-DIO-hChR2(H134R)-eYFP (ChR2-eYFP), 3.2×10^{12} particles ml^{-1} for AAV5/EF1a-DIO-eArch3.0-eYFP (eArch3.0-eYFP), 4.0×10^{12} particles ml^{-1} for AAV5/EF1a-DIO-eNpHR3.0-eYFP (eNpHR3.0-eYFP), 4.0×10^{12} particles ml^{-1} for AAV5/EF1a-DIO-eYFP (eYFP control), and 1×10^{12} to 7×10^{12} particles ml^{-1} for AAV2/EF1a-DIO-hChR2(H134R)-eYFP (for *ex vivo* hippocampal electrophysiology). Virus were subdivided into aliquots stored at -80°C until use.

Stereotactic surgery. Anaesthesia was induced using isoflurane (induction, 5%; maintenance, 1–2%; air-flow, 1 l min^{-1}). The animals were placed in the stereotactic frame (Kopf Instruments). For viral or retrobead (Lumafuor) injection, a small hole was drilled into the skull over the target site. The virus (0.75–1 μl) or retrobead solution (0.1 μl) was then injected at $0.1 \mu\text{l min}^{-1}$ into the target site using a Nanofil syringe (WPI) and UMP3 pump (WPI) mounted directly on the stereotactic frame. After each injection, the needle was kept in place for 10 min to ensure proper diffusion of the virus. Animals recovered on a heating pad until normal behaviour resumed. All experiments involving viral constructs were performed at least 3 weeks after surgery to allow for sufficient expression. Viral infusion coordinates were VTA (from bregma⁵³: anterior–posterior (AP), -3.50 mm ; mediolateral (ML), 0.50 mm ; and dorsal–ventral (DV) from the dura, -4.40 mm) and LC (AP, -5.45 mm ; ML, 1.20 mm ; and DV, -3.65 mm).

Event arena pharmacological experiment (Fig. 1). Bilateral 26-gauge microinjection steel guide cannulae (2.5 mm length, 3.0 mm distance between cannulae; Plastics One) with stylets that protruded 0.5 mm below the end of the cannula (33 gauge, Plastics One) were implanted into the dorsal hippocampus. The cannula implantation coordinates were (AP, -2.10 mm ; ML, $\pm 1.50 \text{ mm}$; and DV, -2.00 mm).

Extracellular recording during novelty exploration (Fig. 2). The Cre-inducible AAV5 ChR2-eYFP virus (1 μl) was unilaterally injected into VTA or LC as mentioned above. Four jeweller's screws were then placed in the skull and the ground wire was attached to one of the skull screws. The microdrive implantation coordinates were VTA (AP, -3.52 mm ; ML, 0.48 mm ; and DV, -4.00 mm) and LC (AP, -5.45 mm ; ML, 1.00 mm ; and DV, -2.80 mm). Adhesive cement (C&B metabond, Parkell) and dental acrylic were then sculpted around the microdrive.

Tract tracing experiment (Fig. 3 and Extended Data Fig. 5). For anterograde tracing, the Cre-inducible AAV5 eYFP virus (1 μl) was unilaterally injected into VTA or LC as described above. For retrograde tracing, retrobeads (0.1 μl) were unilaterally injected into CA1 (AP, -2.18 mm ; ML, 1.18 mm ; and DV, -1.36 mm), CA3 (AP, -2.18 mm ; ML, 2.68 mm ; and DV, -2.05 mm) and DG (AP, -2.18 mm ; ML, 1.36 mm ; and DV, -1.82 mm).

Event arena experiment with optogenetics (Fig. 4). Arterial oxygen saturation, heart rate and breath rate were monitored by MouseOx instrument (STARR Life Science). The Cre-inducible AAV5 virus (ChR2-eYFP or eYFP, 1 μl per side) was injected bilaterally into VTA and LC according to the procedure described above. A dual ferrule optical fibre implant (0.22 numerical aperture, 200 μm core diameter; Doric Lenses) was implanted vertically into VTA (AP, -3.40 mm ; ML, $\pm 0.50 \text{ mm}$; and DV, -4.30 mm). Subsequently, a two-ferrule optical fibre implant (0.22 numerical aperture, 200 μm ; Doric Lenses) was implanted into LC at -30° angle to coronal plane (AP, -5.45 mm ; ML, $\pm 0.90 \text{ mm}$; and DV, -3.00 mm). Additionally, bilateral 26-gauge steel guide cannulae (4.0 mm length, 3.0 mm distance between

cannulae) with stylets (33 gauge) that protruded 0.5 mm below the end of the cannula was inserted into the dorsal hippocampus at a 40° angle to coronal plane (AP, -2.10 mm ; ML, $\pm 1.50 \text{ mm}$; and DV, -2.00 mm).

Event arena experiment with pharmacological inactivation (Fig. 6). The Cre-inducible AAV5 virus (eArch3.0-eYFP or eYFP, 1 μl per side) was injected bilaterally into LC, and a two-ferrule optical fibre implant was then implanted into LC at -30° angle to coronal plane as described above. Additionally, bilateral 26-gauge steel guide cannulae (4.8 mm length, 1.0 mm distance between cannulae) with stylets (33 gauge) that protruded 0.5 mm below the end of the cannula was inserted into VTA (AP, -3.40 mm ; ML, $\pm 0.50 \text{ mm}$; and DV, -4.40 mm).

Ex vivo hippocampal electrophysiology (Fig. 5). The Cre-inducible AAV2 ChR2-eYFP virus (0.75–1 μl per side) was bilaterally injected into LC (AP, -5.45 mm ; ML, $\pm 0.90 \text{ mm}$; and DV, -3.00 mm) over a 10–15 min period using a borosilicate glass electrode (10–15 M Ω) pulled with a horizontal pipet puller (P-97, Sutter Instrument) and a picospritzer (Parker) timed by a Master-8 pulse stimulator (A.M.P.I.). After each injection, the pipette was kept in place for 5 min to ensure proper diffusion of the virus.

Postoperative analgesia. Carprofen (0.08 ml kg^{-1} body weight), or buprenorphine (0.1 mg kg^{-1} body weight), was administered by subcutaneous injection at the end of all surgical procedures. All mice were allowed a recovery period of at least 7 days for them to regain their pre-surgery weights before electrophysiological or behavioural testing.

Everyday memory apparatus. Everyday memory was tested in an event arena: a square open field (120 cm wide \times 120 cm long) with walls (35 cm high) made out of transparent Plexiglas (Fig. 1a) with four adjacent start boxes (black Plexiglas). The name 'event arena' derives from it being an arena in which 'events' happen (for example, finding food)¹³. The floor of the arena, arranged in a 5×5 grid, was covered with $\sim 2 \text{ cm}$ of sawdust and had two intramaze landmarks (a white metal cube located at row 3, column 2, and a black rubber flash light at row 3, column 4). The Plexiglas sandwells, in which food reward was potentially available, could be fitted into any of the 23 remaining sandwell positions (positions occupied by internal cues were excluded). The mice had access to the arena and sandwells when the startbox door was opened in any trials. Light levels, checked every day, were 25–35 lx. Data were recorded using custom-made LabVIEW software (National Instruments), using the image from camera placed above the arena.

Novelty apparatus. For novelty exploration, a square Plexiglas open field with transparent walls (70 cm wide \times 70 cm long \times 30 cm high) was placed in the middle of the event arena (Fig. 1b). To maintain the novelty of the environment, a wide range of floor substrates (dried leaves, shredded paper, feathers, acrylic pompoms, corks, lolly sticks, Lego blocks, pipe cleaners, shredded straws and sea shells) that covered the floor of the box were used, as in a previously published study that used rats¹³.

Event arena tasks. Sample size, randomization, blinding and replication. Sample size was determined based on variability in pilot data. A distinctive feature of event arena tasks is that most but not all comparisons are 'within-subject' design in which every single subject is exposed to every single treatment, including the control treatment. This typically reduces the number of subjects required for statistical significance and avoids issues associated with randomization. All non-rewarded probe tests (see below) were analysed blind and, being conducted against a stable performance background, were typically conducted twice or three times (internal replication). Averaging data helped reduce variability.

Shaping and habituation. After handling, habituation (eight sessions) involved training mice to dig in the sandwells to retrieve food (a half of cereal 'Cheerios') and carry it to the start box.

Everyday training protocol. The goal in each daily session was to encode the changing daily location of a rewarded sandwell encountered during two consecutive sample trials (two retrievals of buried food in each trial), and then, 10 min later, return to that same location during the choice trial (Fig. 1a). The choice trial was a retrieval test that involved a rewarded sandwell in a location that matched the sample location (the 'correct' location; win-stay rule) and four non-rewarded sandwells placed in other locations around the arena (the 'incorrect' locations). Training sessions were conducted daily (5–7 sessions per week) using 16 different sandwell configurations with rewarded sandwell positions counterbalanced between mice (Extended Data Fig. 1a). We calculated a performance index, using the formula performance index = $100 - [100 \times (\text{errors}/4)]$, based on the number of errors made during the choice trial (an error being defined as digging at an incorrect sandwell). The value expected on the basis of chance was 50% (two errors). With each behavioural cohort, we began conducting critical memory probe tests once mice reached average performance index of 75% (equivalent to one error, computed as average of last five training sessions). This typically happened within 35 training sessions, but sometimes additional training was needed after surgical procedures.

Memory probe tests. The primary data measures of the study were derived from 'memory tests' performed as 'probe' tests, defined as sessions in which none of the sandwells contained any accessible food pellets. The mice were cued with a food pellet in the start box, and then allowed to search for the correct sandwell for 60 s from the first dig of any sandwell. After 60 s, the experimenter quietly entered the room and buried pellets in the correct sandwell, allowing the mouse to retrieve them (one by one as in training, this limited extinction). Dig time at each sandwells was measured, and the relative proportion of time at the correct and incorrect sandwells was calculated. The value expected on the basis of chance was 20%. Probe tests were always separated by at least two sessions of regular training. In the 'reward magnitude' probe tests (Fig. 1b and Extended Data Fig. 1d), animals retrieved either two pellets (low reward) or eight pellets (high reward) during memory encoding.

Novelty exploration. The mice underwent seven sessions of habituation to the box with sawdust placed in the event arena. For post-encoding unexpected environmental novelty, the mouse was placed into the centre of the box lined with a novel floor surface and allowed to explore freely for 5 min.

Behavioural pharmacology. Microinfusion/injection of drugs. To help reduce any stress, all drugs were infused in the home cages. The stylets in the guide cannulae were replaced by a double infusion cannula (33 gauge, Plastics One) connected to two 5- μ l microsyringes (WPI) in a microinfusion pump (Native Instruments) via flexible plastic tubing (C232CS, Plastics One) filled with Fluorinert (3 M). The tips of infusion cannulae projected 0.5 mm below the tip of the guide cannulae. For intra-hippocampal microinjection, 0.5 μ l of drug per cannula was infused at 0.2 μ l min⁻¹ (2.5 min). Infusion cannulas were left in place for a further 2.5 min before being replaced with stylets to aid drug absorption. For intra-VTA microinjection, 0.3 μ l was injected at a rate of 0.3 μ l min⁻¹ (1 min) followed by 1 min of waiting. The mice were habituated to the experimental procedure of injection and to vehicle injection before the drug test to minimize the potential novelty of the procedure. Mice received drug injection 20 min (hippocampal microinfusions and intraperitoneal injection of clonidine) or 3 min (VTA microinfusions) before the novelty exploration.

Drug concentrations. For microinfusions, the concentrations used were 21.1 mM (6.25 μ g μ l⁻¹) for β -adrenoceptor antagonist propranolol ((S)-(-)-propranolol hydrochloride, 295.80 g mol⁻¹; Sigma-Aldrich), 3.1 mM (1 μ g μ l⁻¹) for D₁/D₅ receptor antagonist SCH23390 (SCH 23390 hydrochloride, 324.24 g mol⁻¹; Tocris) and 2% w/v for voltage-gated sodium channel blocker lidocaine (lidocaine hydrochloride monohydrate, 288.81 g mol⁻¹; Sigma-Aldrich). α_2 -Adrenoceptor agonist clonidine (clonidine hydrochloride, 266.55 g mol⁻¹; Sigma-Aldrich) was administered intraperitoneally at a dose of 50 μ g kg⁻¹ of body weight. We used 0.9% NaCl in H₂O (saline) as a vehicle and for control infusions. Both vehicle and drug solutions were stored in 100 μ l aliquots at -20 °C until use.

Optogenetic photostimulation in the event arena experiment. The mice were extensively habituated to the experimental procedure of photostimulation and to flickering blue light for several weeks before the optogenetic photostimulation test to minimize the potential novelty of the procedure. Laser stimulation, consisting of 20 5-ms pulses of 473-nm light at 25 Hz, delivered every 5 s (average stimulation rate 4 Hz) for the duration of 5 min (Fig. 4b), was performed in home cages using two blue solid-state diode pumped lasers (18–19 mW, Laser 2000) connected to either a dual fibre optic patch cord (for VTA; 0.22 numerical aperture, 200 μ m core diameter; Doric Lenses) or two single fibre optic patch cords (for LC; 0.22 numerical aperture, 200 μ m, Doric Lenses). Both lasers were synchronously controlled using custom-built LabVIEW software.

Extracellular recording during novelty exploration. Apparatus and light stimulation. The behavioural apparatus consisted of a rectangular wooden box with three compartments: a screening chamber with space for the home cage, as well as 'familiar' and 'novel' chambers (both 30 cm wide \times 70 cm long). The apparatus was surrounded by black curtains and light level was kept at 25–35 lx. The floor of the familiar environment was covered in sawdust (the floor substrate of the event arena), and the floor of the novel environment was covered in one of the floor substrates used as novelty in the event arena experiments. Each floor substrate was used only once for each mouse. Unit activity was recorded extracellularly using the implanted custom-built screw-driven microdrive consisting of a 200 μ m optic fibre surrounded by four tetrodes (an 'optetrode') that protruded 400–800 μ m beyond the fibre tip⁵⁴. Signals were fed through a 16 channel unity gain headstage amplifier (Axona), band-pass filtered at 300–5,000 Hz, amplified 1,000–40,000 times, digitized at 50 kHz and stored for subsequent analysis. Spike capturing was done on-line using amplitude threshold. Recorded neurons were identified as TH⁺ using Cre-dependent ChR2 expression and low-frequency light stimulation⁵⁵. Laser stimulation was performed using a blue solid-state diode pumped laser (473 nm, Laser 2000) connected to a fibre optic patch cord (0.22 numerical aperture, 200 μ m core diameter), and controlled with the data capturing software (Axona). Epochs of

60 light pulses (1 Hz, 5 ms pulse duration) at different light intensities (0.1–20 mW) were then administered and each tetrode was screened for light-evoked spikes. Spikes were classified as 'light-evoked' if their latency from the onset of the light pulse was between 0 and 15 ms, all other spikes being classified as 'spontaneous'. Units were classified as light-responsive if (a) a cell fired a light-evoked spike in response to more than one third of light pulses, (b) the shape of the mean light-evoked waveform of a unit closely resembled the spontaneous waveform of the same unit. Units with basal firing rates above 20 Hz were excluded from this analysis because of intrinsically high probability of spiking within the 15 ms window after the light pulse.

Novelty exploration. Two weeks after surgery, the implanted mice underwent 5 days of habituation to the experimental apparatus and the familiar environment. Following habituation, mice underwent daily screening trials for light-responsive neurons in the home cage. If no light-responsive cells were found, the mouse was allowed to explore the familiar environment for 5 min and was then unplugged. If one or more recording channels showed a light-responsive unit, the mouse was subjected to an 'exploration' trial after a 10 min delay (Fig. 2b). The mouse might first be placed in the novel environment for 5 min, followed by 25 min in the home cage and 5 min in the familiar environment; alternatively, the order of novel and familiar exploration was reversed (in a counterbalanced manner). Baseline recording (5 min) was performed in the home cage between two exploration trials. These particular time delays were chosen to mimic timing used in the event arena experiments, where there was a 30 min delay between memory encoding and novelty exploration. The light-sensitive unit was again confirmed in home cage with 1 Hz light stimulation after open field exploration. The microdrive was advanced by \sim 40 μ m at the end of daily session to ensure that recordings are made from a different population of neurons.

Recording and analysis. Recorded spikes were clustered using Klusterkwik 1.5 unsupervised clustering algorithm (<http://klusta-team.github.io/klustakwik/>) on the basis of their energy and first principal component of the waveform. Clusters were then corrected manually using Klusters spike sorting software (<http://neurosuite.sourceforge.net/>), on the basis of several additional parameters (width of waveform, amplitude, time at peak, auto- and cross-correlograms). Data were analysed using Matlab R2012a (MathWorks). Firing patterns were characterized in terms of firing rate, rate of burst events and firing rate of spikes within bursts. Bursts were defined, using classic criteria⁵⁶, as trains of two or more spikes with an interspike interval of less than 80 ms, followed by an interspike interval of more than 160 ms. For comparison of novelty modulation in VTA-TH⁺ and LC-TH⁺ neurons, firing rates of individual neurons in the novel and familiar environments were binned in 10-s bins and normalized to the average home cage firing rate of all identified neurons in respective brain areas. For additional analysis of the novelty modulation, binned firing rates of individual neurons in the novel environment were z-scored to their respective firing rates in the familiar environment.

Anatomical tract tracing. Under deep pentobarbital anaesthesia (100 mg kg⁻¹ of body weight, intraperitoneally), the mice were fixed transcardially with 4% paraformaldehyde in 0.1 M sodium phosphate buffer, pH 7.2, post-fixed in the same fixative for 24 h, and placed in 30% sucrose in phosphate buffer. Sections of fixed brains (30 μ m in thickness) were prepared using a freezing microtome (SM2000R, Leica Microsystems) for immunohistochemistry.

Immunofluorescence. Antibodies used included: goat anti-GFP⁵⁷, guinea-pig anti-NET⁵⁸ and mouse anti-TH (AB152, Millipore). eYFP was visualized by anti-GFP immunostaining. All immunohistochemical incubations were done at room temperature. Sections were incubated successively with 10% normal donkey serum for 20 min, a mixture of primary antibodies overnight (1 μ g ml⁻¹), and a mixture of Alexa Fluor 488-, Cy3- or Cy5-labelled species-specific secondary antibodies (Invitrogen; Jackson ImmunoResearch) for 2 h at a dilution of 1:200.

Fluorescent *in situ* hybridization. Brains were freshly obtained under deep diethyl ether anaesthesia and immediately frozen in powdered dry ice. Fresh-frozen sections (20 μ m) were cut on a cryostat (CM1900, Leica Microsystems). All sections were mounted on silane-coated glass slides. Mouse cDNA fragments of TH (bases 1–1025; GenBank accession number AY855842), and NET (bases 124–814, MMU76306) were subcloned into the pBluescript II plasmid vector. Digoxigenin- or fluorescein-labelled cRNA probes were transcribed *in vitro* for fluorescent *in situ* hybridization⁵⁹.

Image acquisition and data analysis. Images were taken with a confocal laser-scanning microscope (FV1200, Olympus) equipped with diode laser lines, and UPlanSApo (20 \times , numerical aperture 0.75) and PlanApoN (60 \times , numerical aperture 1.4, oil-immersion) objective lenses (Olympus). To avoid cross talk between multiple fluorophores, Alexa Fluor 488, Cy3, and Alexa Fluor 647 fluorescent signals were acquired sequentially using the 473, 559, and 647 nm excitation laser lines, respectively. All images show single optical sections. For quantification of anterograde tracing, we obtained images with a 20 \times objective and then

created images of the entire hippocampus with a Fluoview image stitching software (Olympus). For analysis, the separate colour components were converted to greyscale, and the area of eYFP- and TH-positive elements were measured with Integrated Morphometry Analysis module (MetaMorph software, Molecular Devices).

Ex vivo hippocampal electrophysiology. Hippocampal slice preparation. Thick coronal slices (300 μm) containing the hippocampus were cut from Th-Cre transgenic mice³² expressing AAV2 ChR2-eYFP in LC (8–12 weeks old) in low light conditions to prevent unwanted ChR2 activation. Animals were anaesthetized under 1.5–2% isoflurane, and the brains removed and blocked following rapid decapitation. Hippocampal slices were prepared using a vibratome (VT 1000S, Leica Microsystems) in ice cold *N*-methyl-D-Glucosamine (NMDG) ringer solution (in mM): 5 NaCl, 57 NMDG, 37.5 Na-pyruvate, 12.5 Na-lactate, 5 Na-ascorbate, 2.5 KCl, 1.25 NaH_2PO_4 , 25 NaHCO_3 , 25 glucose, 10 $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.5 $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, the pH was set between 7.3 and 7.4 using 12 N HCl, the osmolarity was adjusted as needed to 315 mOsm using glucose and the solution was bubbled with 95% O_2 and 5% CO_2 gas. Slices were maintained in NMDG ringer at room temperature for no longer than 15 min and then transferred to artificial cerebrospinal fluid (in mM): 125 NaCl, 2.5 KCl, 1.25 NaH_2PO_4 , 1.3 MgCl_2 , 2 CaCl_2 , 25 NaHCO_3 and 25 dextrose continuously bubbled with 95% O_2 and 5% CO_2 gas, where they were kept up to 6 h, protected from light, for experimentation. One slice per animal was used.

Ex vivo whole-cell recordings. Slices were transferred to a submersion recording chamber and were perfused with artificial cerebrospinal fluid at a rate of 1–2 ml min^{-1} at 26–29°C. EPSC recordings were performed with GABA_A receptor antagonist picrotoxin (602.58 g mol^{-1} ; Sigma-Aldrich) at a concentration of 50 μM and D₂ receptor antagonist eticlopride (eticlopride hydrochloride, 377.31 g mol^{-1} ; Sigma-Aldrich) at a concentration of 100 nM in the bath. A borosilicate glass electrode (3–5 M Ω), pulled with a horizontal pipet puller (P-97), was filled with Cs-methanesulfonate pipet solution (in mM): 110 CsMeSO₃, 15 CsCl, 8 NaCl, 2 EGTA, 10 HEPES, 3 QX-314, 2 ATP and 0.3 GTP adjusted to 295 mOsm and pH 7.3. Whole-cell pyramidal cell recordings from area CA1 were acquired using a combination of visualized and blind patch techniques. Cells were held at –60 mV using a multiclamp 700B amplifier (Molecular Devices). A bipolar stimulating electrode (FHC) was placed in the stratum radiatum region of dorsal CA1 within 100–200 μm of the recording electrode and stimulation (delivered at a rate of 0.2 Hz) was set to elicit current responses of 50–150 pA. Data were acquired and analysed automatically using P-Clamp 10 (Molecular Devices). Recordings were discarded if the series resistance varied by more than 20% or if the initial holding current exceeded 70 pA. Following a 5 min baseline acquisition, hippocampal slices were given a 470-nm light stimulus (consisting of three trains delivered every 2 min, of 60 5-ms pulses, applied at 18 Hz (Extended Data Fig. 7b)) through the 40 \times objective lenses. Following the light stimulus, baseline stimulation resumed.

Ex vivo field recordings. Slices were transferred to a submersion chamber and were perfused with artificial cerebrospinal fluid at a rate of ~2 ml min^{-1} at 29–31°C. Field recordings from the stratum radiatum of dorsal CA1 were acquired using a borosilicate glass electrode (1–3 M Ω) filled with artificial cerebrospinal fluid. A bipolar stimulating electrode was also placed in the stratum radiatum of CA1 within 100–200 μm of the recording electrode and stimulation (one stimulus every 30 s) was set to elicit a fEPSP slope that was ~50% of the maximum value. A stable 15 min baseline was obtained, followed by 10 more min of baseline stimulation with or without simultaneous optogenetic stimulation of LC-TH⁺ axons. Photostimulation consisting of four trains, at a 1 s interval, consisting of four 10-ms pulses of 470-nm light at 16 Hz, delivered every 30 s for the duration of 10 min (Extended Data Fig. 7c), was applied through the 10 \times objective (directly before the Schaffer collateral stimulus). After the 25 min baseline, a weak theta-burst tetanus was applied consisting of four trains, at a 100 ms interval, consisting of three pulses at 50 Hz (12 pulses in total). Baseline stimulation then resumed as described above for 45 min. Every two traces were averaged to reduce variability. Data were acquired and analysed automatically using P-Clamp 10.

Ex vivo pharmacology. Where indicated, the following drugs were bath applied: α -adrenoceptor antagonist prazosin (prazosin hydrochloride, 419.86 g mol^{-1} ; Sigma-Aldrich) at a concentration of 30 μM , β -adrenoceptor antagonist propranolol ((S)-(-)-propranolol hydrochloride, 295.80 g mol^{-1} ; Sigma-Aldrich) at a concentration of 30 μM , D₁/D₅ receptor antagonist SCH 39166 (SCH 39166 hydrochloride, 394.73 g mol^{-1} ; Tocris) at a concentration of 100 nM (intracellular recordings, $n = 2$) and D₁/D₅ receptor antagonist SCH 23390 (SCH 23390 hydrochloride, 324.24 g mol^{-1} ; Tocris) at a concentration of 100 nM (intracellular recordings, $n = 3$) or 1 μM (extracellular recordings).

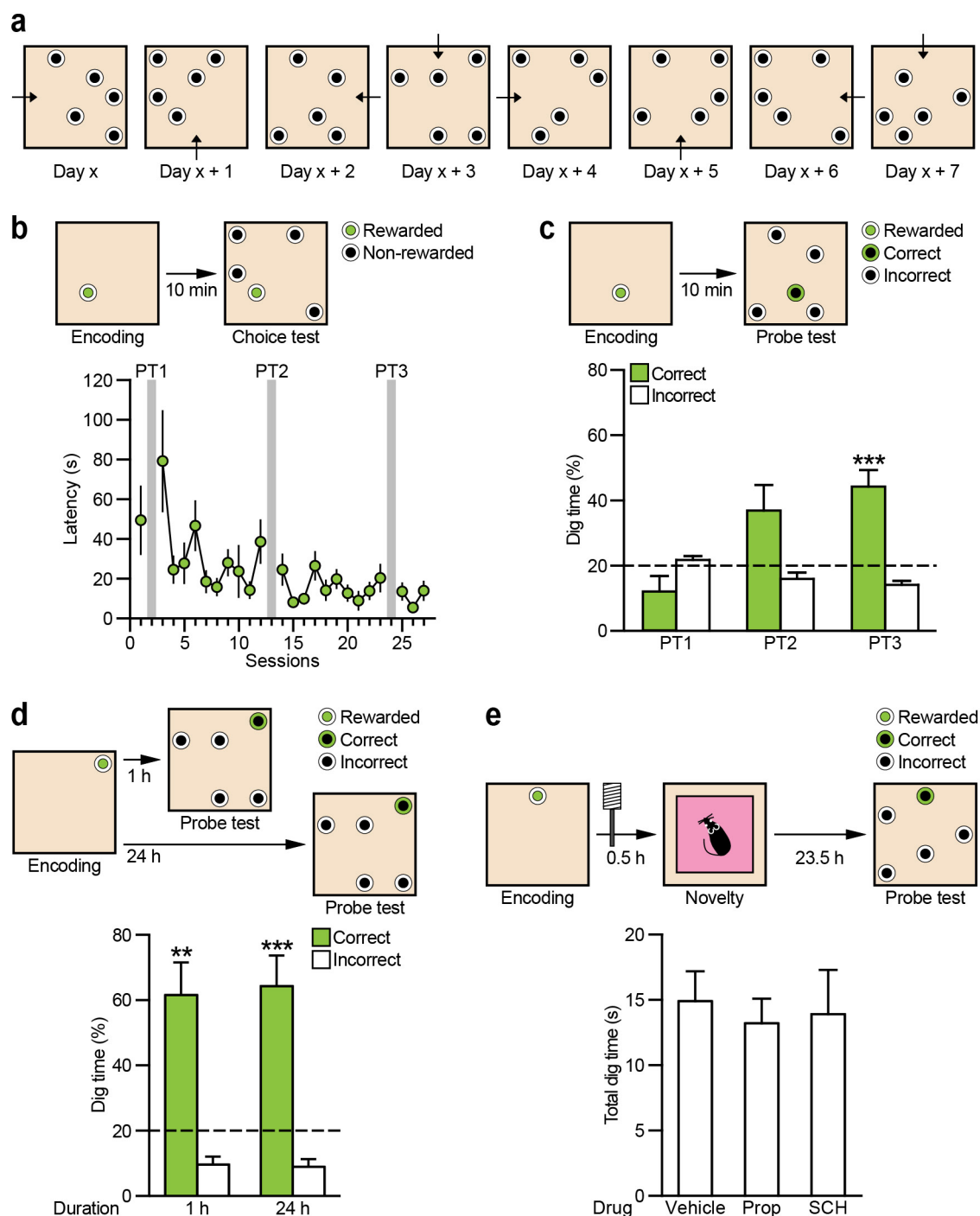
Acute extracellular recordings in anaesthetized mice. Anaesthesia was induced using isoflurane for LC optrode recording as described above or with 4.8% chloral hydrate for VTA recording (induction: 480 mg kg^{-1} body weight, intraperitoneally; maintenance: 120 mg kg^{-1} body weight, intraperitoneally). Recordings were made using a 125 μm 1 M Ω tungsten electrode (A-M systems). This electrode was connected to a differential AC amplifier (A-M Systems), signals were band-pass filtered at 300 Hz to 5 kHz, amplified 10,000 times and digitized at 20 kHz. Spiking activity was defined as spikes that exceed five standard deviations from the mean value of the baseline signal (1 min before laser stimulation (LC) or drug infusion (VTA)). Multi-unit activity in each trace was then normalized to the pre-stimulation (LC) or pre-infusion (VTA) baseline.

LC optrode recording. The tungsten electrode was coupled to an optic fibre (0.22 numerical aperture, 200 μm core diameter) (an 'optrode'). This optrode was positioned above LC (AP, –5.45 mm; ML, 1.00 mm; and DV, –2.80 mm), and was then gradually lowered in 50 μm increments until multi-unit activity was observed. Laser stimulation was performed using a green solid-state diode pumped laser (532 nm, Laser 2000) with 10–20 mW output from the fibre. For quantification of eArch3.0 recordings, baseline was measured over 30 s before the start of illumination (Pre), level of inhibition was measured over 5 min light on period (LC-on), rebound activity was measured over 1 min after the end of illumination (rebound) and post-inhibition baseline was measured 4–5 min after the end of illumination (Post).

VTA recording with pharmacology. The drug cannula (33 gauge, Plastics One) was positioned in the VTA at 14° angle to sagittal plane (AP, –3.52 mm; ML, 0.48 mm; and DV, –4.40 mm) and the recording electrode was positioned vertically at the boundary of the VTA. For intra-VTA microinjection of lidocaine, infusion parameters were the same as those used in the behavioural experiment as described above. For quantification, pre-infusion baseline was measured over 30 s before the start of infusion (Pre), level of inhibition was measured 3–8 min after the start of infusion (Lid)—the period that corresponds to novelty exploration, and post-inhibition baseline was measured 17–18 min after the start of injection (Post).

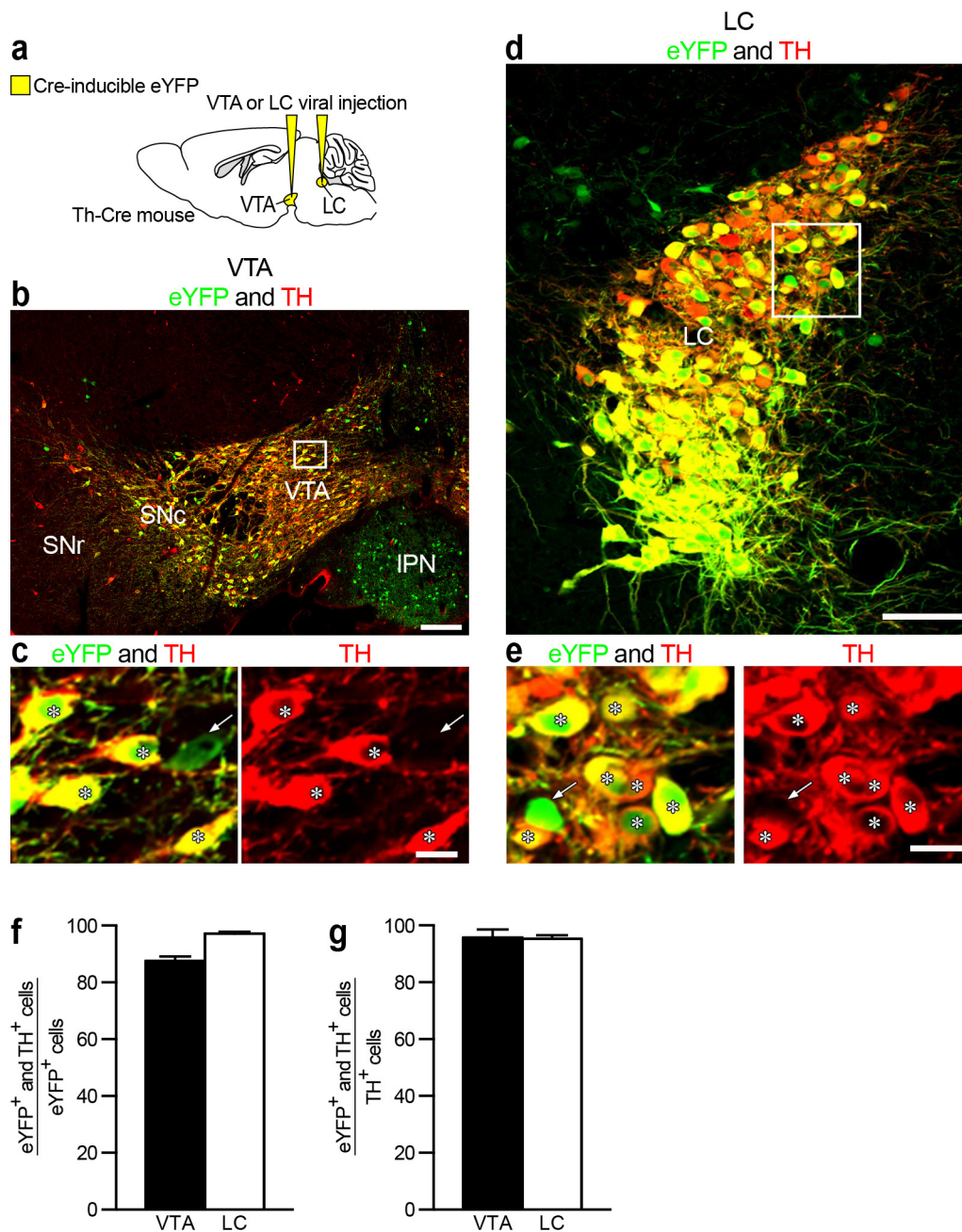
Statistics, data presentation and data deposition. Statistical analyses were performed using SPSS version 19 (IBM). All data are expressed as mean \pm s.e.m. Statistical significance was always determined by ANOVAs, before orthogonal comparisons where possible or Tukey's HSD tests as appropriate to correct for multiple comparisons, paired *t*-tests and one-sample *t*-tests. All statistical tests were two-tailed. Analysis of probe test performance was done on the basis on the '% correct dig time' score. In the pharmacological inactivation experiment (Fig. 6), three animals that persistently failed to show any novelty-induced memory enhancement in all control conditions (all 24-h probe test scores (% correct dig) in the 'novelty with vehicle' condition at below chance level, that is, <20%) were eliminated from the whole data set before statistical analysis (leaving $n = 15$). The rationale for this was that it was not possible to measure the impact of pharmacological manipulations on the novelty effect in animals that are not susceptible to it, the existence of the novelty effect having been established in the first cohort of mice (Fig. 1b). All source data for the preparation of graphs and statistical analysis are presented online. All other relevant data that support the conclusions of the study are available from the authors on request.

- Lindeberg, J. *et al.* Transgenic expression of Cre recombinase from the tyrosine hydroxylase locus. *Genesis* **40**, 67–73 (2004).
- Savitt, J. M., Jang, S. S., Mu, W., Dawson, V. L. & Dawson, T. M. Bcl-x is required for proper development of the mouse substantia nigra. *J. Neurosci.* **25**, 6721–6728 (2005).
- Franklin, K. B. J. & Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates* 3rd edn (Academic, 2007).
- Anikeeva, P. *et al.* Optrode: a multichannel readout for optogenetic control in freely moving mice. *Nature Neurosci.* **15**, 163–170 (2011).
- Cohen, J. Y., Haesler, S., Vong, L., Lowell, B. B. & Uchida, N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* **482**, 85–88 (2012).
- Grace, A. A. & Bunney, B. S. Intracellular and extracellular electrophysiology of nigral dopaminergic neurons—1. Identification and characterization. *Neuroscience* **10**, 301–315 (1983).
- Takasaki, C. *et al.* Cytochemical and cytological properties of perineuronal oligodendrocytes in the mouse cortex. *Eur. J. Neurosci.* **32**, 1326–1336 (2010).
- Uchigashima, M. *et al.* Subcellular arrangement of molecules for 2-arachidonoyl-glycerol-mediated retrograde signaling and its physiological contribution to synaptic modulation in the striatum. *J. Neurosci.* **27**, 3663–3676 (2007).
- Yamasaki, M., Matsui, M. & Watanabe, M. Preferential localization of muscarinic M1 receptor on dendritic shaft and spine of cortical pyramidal cells and its anatomical evidence for volume transmission. *J. Neurosci.* **30**, 4408–4418 (2010).



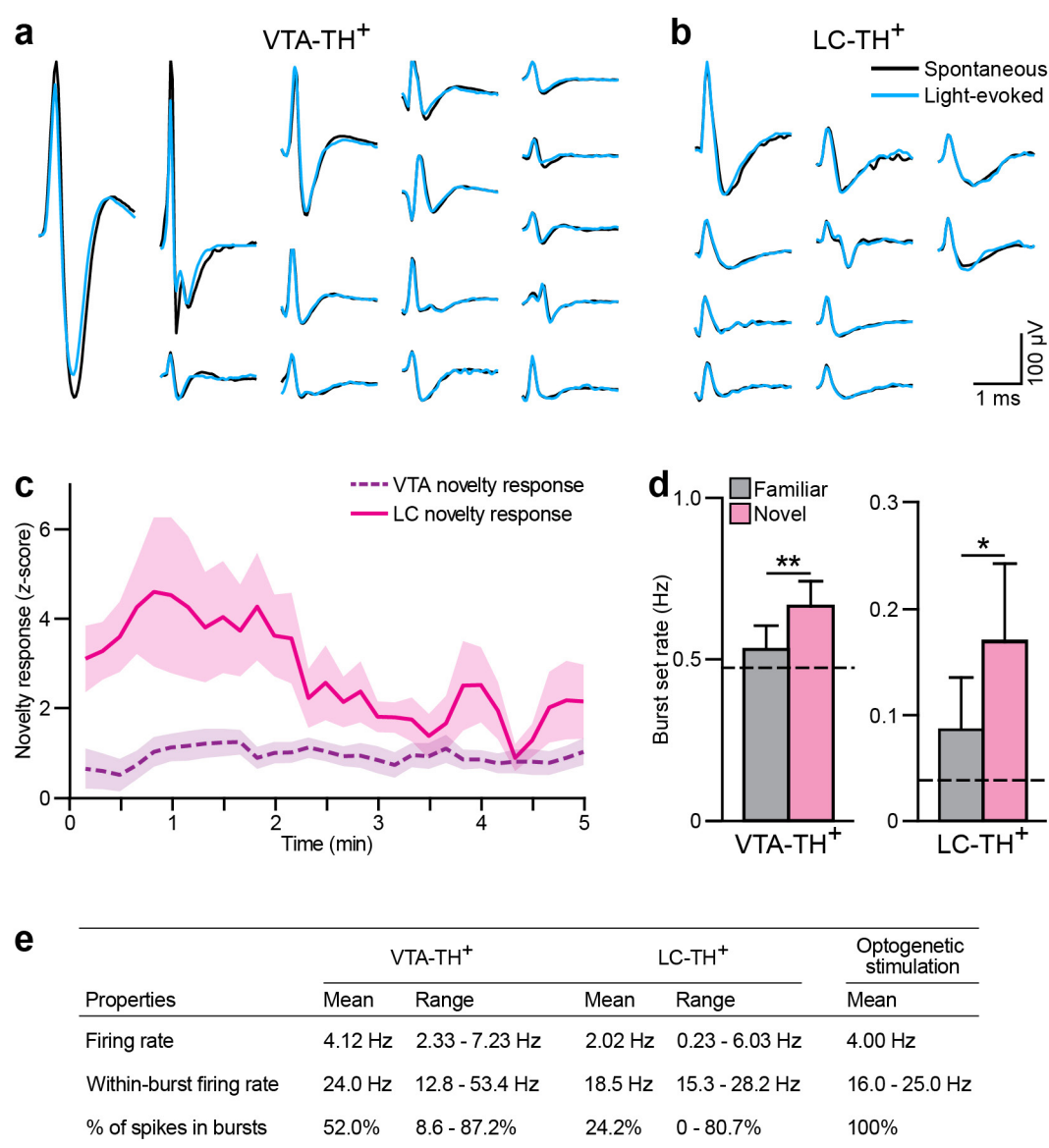
Extended Data Figure 1 | Everyday spatial memory task in the event arena. **a**, Example sandwell locations (black circles) and starting positions (black arrows) used during regular training and non-rewarded probe tests (probe tests). Sixteen different sandwell configurations were used throughout experiments, with daily rewarded sandwell positions counterbalanced between animals. **b**, Latency to dig in correct sandwell during choice test decreased as initial training progressed (Fig. 1a; Th-Cre mice, $n = 13$). Grey bars mark acquisition probe test sessions. **c**, Ten-minute probe tests conducted at different stages of initial training illustrate the learning curve ($F_{2,24} = 9.35$, $P < 0.001$), with animals performing

above chance on probe test 3 (one-sample t -test versus chance, $t_{12} = 4.76$, $P < 0.001$). **d**, Increased reward during the sample trial resulted in persistent memory at 24 h (1 h versus 24 h: $t_{12} < 1$; 1 h: $t_{12} = 4.16$, $P < 0.01$; 24 h: $t_{12} = 4.70$, $P < 0.001$). **e**, Impact of intra-hippocampal infusion of propranolol (Prop) or SCH23390 (SCH) on animals' performance in 24-h probe test (Fig. 1c). Infusion of Prop or SCH in dorsal hippocampus 20 min before novelty exploration had no impact on the total time that mice spent digging in sandwells ($F_{2,24} < 1$). ** $P < 0.01$, *** $P < 0.001$ versus chance. Dashed line indicates the chance level. Data are means \pm s.e.m.



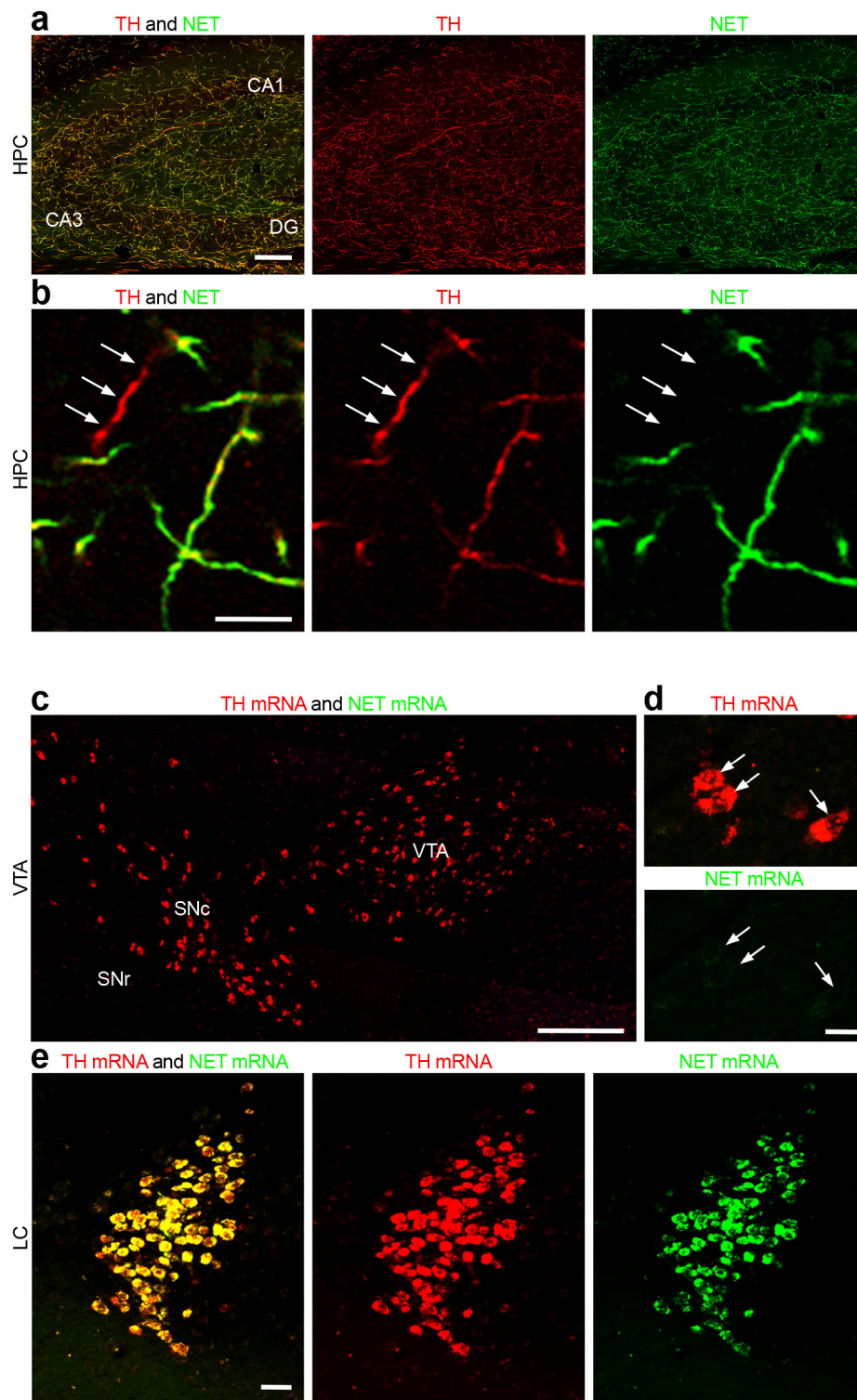
Extended Data Figure 2 | Specificity and expression rate of Cre-inducible AAV in VTA and LC of Th-Cre mice. **a**, Schematic of viral injection. **b–e**, Double immunofluorescence for eYFP (green) and TH (red). In VTA (**b**, **c**) and LC (**d**, **e**), most eYFP-expressing cells are positive for TH (asterisks), and eYFP expression in TH[−] cells (arrows) is only occasionally observed. **f**, **g**, Quantification of specificity (**f**) and expression rate (**g**) of the Cre-inducible AAV in VTA and LC of Th-Cre mice. VTA-TH⁺ neurons are defined as TH⁺ cells located in the parabrachial pigmented area (PBP), paranigral (PN), interfascicular (IF) and rostral linear (RLi) nuclei. Measurements were made from the VTA in the left hemisphere, and three coronal sections (AP: −2.9, −3.5 and −3.9 from

Bregma) from each mouse. LC-TH⁺ neurons are defined as TH⁺ cells located in the lateral floor of the fourth ventricle. Measurements were made from the LC in the left hemisphere, and five coronal sections (AP: −5.3 to −5.7 from Bregma) from each mouse. Total numbers of neurons measured were 2,500 and 1,520 in VTA and LC, respectively. Data are presented as mean ± s.e.m. ($n = 3$ mice in each). IPN, interpeduncular nucleus; LC, locus coeruleus; SNc, Substantia nigra pars compacta; SNr, Substantia nigra pars reticulata; VTA, ventral tegmental area. Scale bars: 200 μ m (**b**, **d**); 20 μ m (**c**, **e**). The mouse brain in this figure has been reproduced with permission from Franklin, K. B. J. & Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates* 3rd ed, 691 (Academic, 2007).



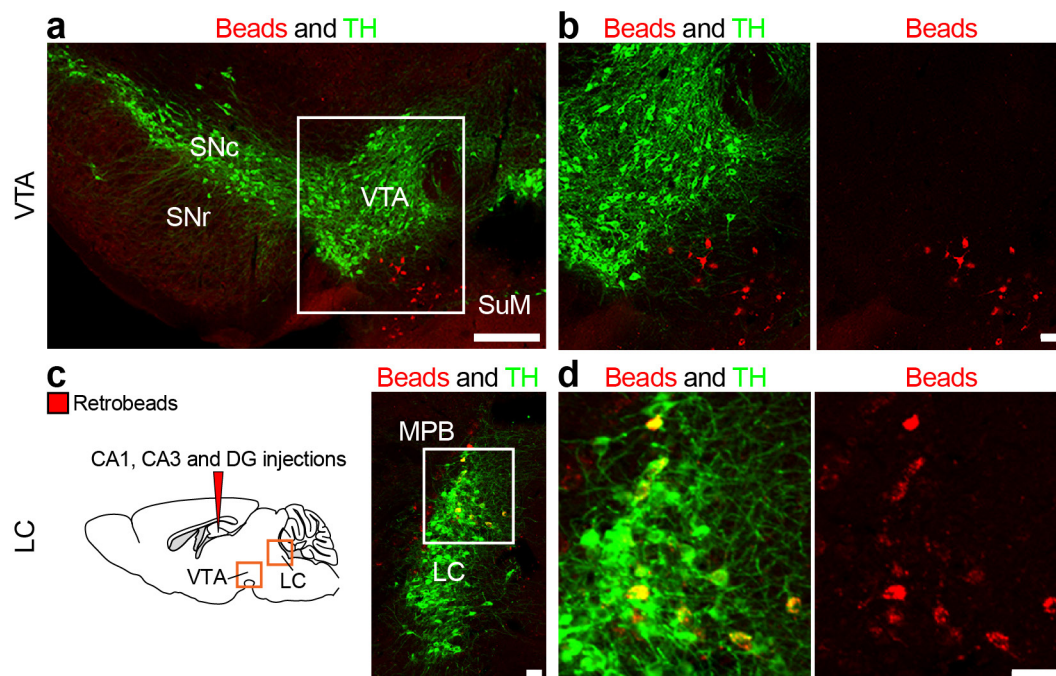
Extended Data Figure 3 | Firing properties of VTA-TH⁺ and LC-TH⁺ neurons. **a, b**, Average spontaneous (black) and light-evoked (blue) waveforms of all identified VTA-TH⁺ neurons in Th-Cre mice expressing ChR2-eYFP in VTA (**a**, $n = 15$ neurons from 5 mice) and LC-TH⁺ neurons in Th-Cre mice expressing ChR2-eYFP in LC (**b**, $n = 10$ neurons from 3 mice) show nearly perfect overlap. **c**, Comparison of the novelty response dynamics of VTA-TH⁺ and LC-TH⁺ neurons. When z-scored to their average firing rates in the familiar environment, LC-TH⁺ neurons showed stronger modulation by novelty than VTA-TH⁺ neurons (two-way ANOVA: brain area \times time interaction, $F_{29,667} = 2.60$, $P < 0.001$). Additionally, LC-TH⁺ neurons but not VTA-TH⁺ neurons displayed habituation in a manner consistent with a novelty signal (one-way ANOVA: main effect of time for LC novel, $F_{29,261} = 2.04$, $P < 0.01$;

no effect of time for VTA novel: $F_{29,406} = 1.18$, $P > 0.05$). **d**, VTA-TH⁺ and LC-TH⁺ neurons fire more bursts in novel than in familiar environments (5 min of exploration) (VTA-TH⁺: two-tailed paired t -test, $t_{14} = 3.70$, $P < 0.01$; LC-TH⁺: $t_9 = 2.48$, $P < 0.05$). Dashed line indicates the baseline burst set rate. **e**, Firing properties of VTA-TH⁺ and LC-TH⁺ neurons in novel environments used to choose physiologically relevant optogenetic stimulation protocols. The average firing rate and within-burst firing rate of our optogenetic stimulation protocols for both *in vivo* and extracellular *ex vivo* experiments are within the physiological range of both VTA-TH⁺ and LC-TH⁺ neurons. In addition, several recorded neurons fired most of their spikes in bursts. * $P < 0.05$, ** $P < 0.01$, paired t -test. Data are means \pm s.e.m.



Extended Data Figure 4 | NET is specifically expressed in LC-TH⁺ neurons. **a**, Immunofluorescence showing overall distribution of TH (red) and NET (green) immunoreactivity in the mouse hippocampus. **b**, Representative high magnification images of double immunofluorescence for TH (red) and NET (green). Note that most TH⁺ axons are co-labelled for NET, and TH⁺-NET⁻ axons (arrows) are only occasionally observed. **c-e**, Double-labelling fluorescence *in situ* hybridization showing distinct

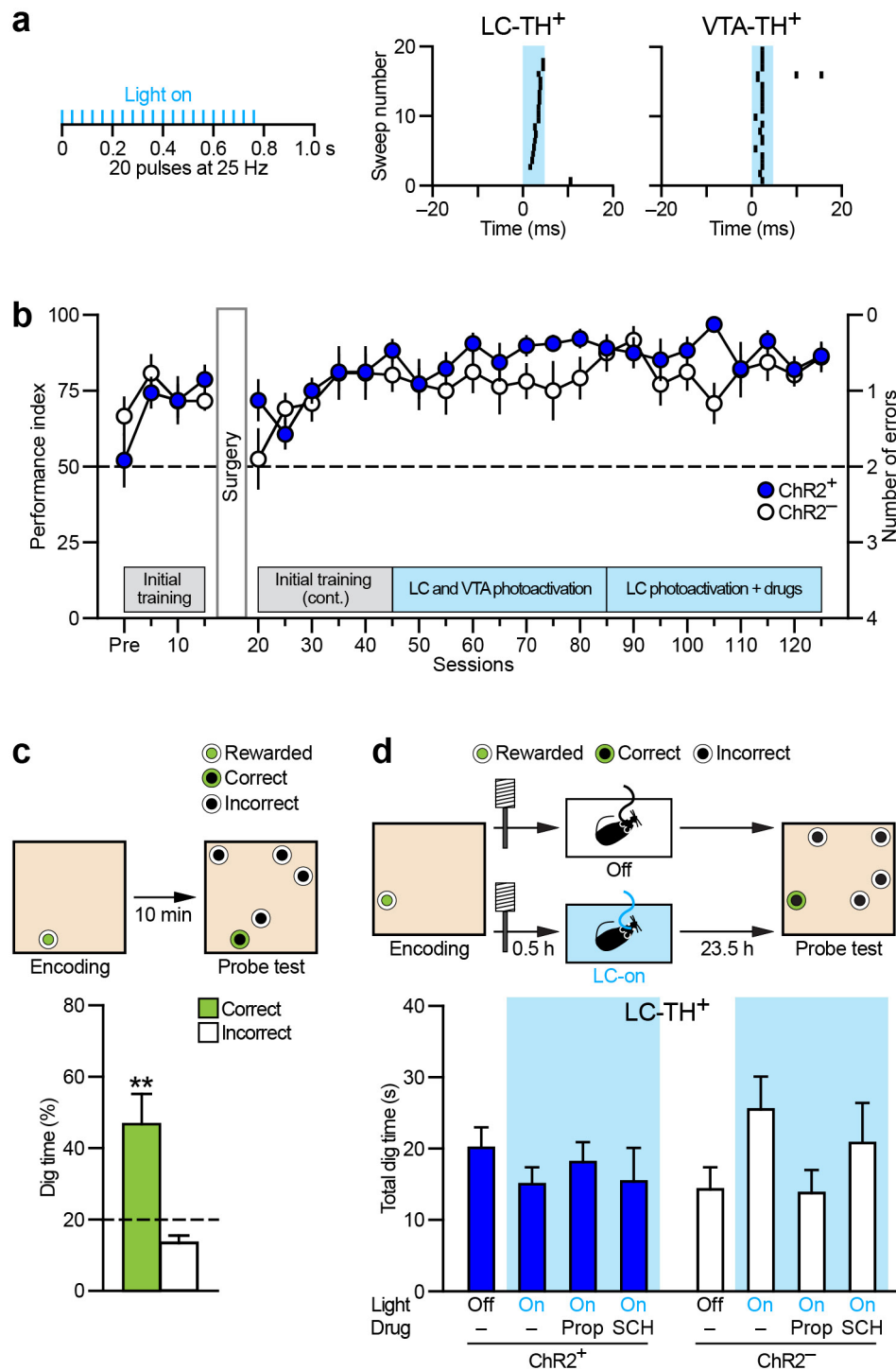
expression pattern of TH (red) and NET (green) mRNAs in VTA (**c, d**) and LC (**e**). Note that mRNA for NET is detected in virtually all TH⁺ neurons in LC, whereas it is not in any of TH⁺ neurons in VTA. CA1 and CA3, hippocampal subregions CA1 and CA3; DG, dentate gyrus; HPC, hippocampus; LC, locus coeruleus; SNc, Substantia nigra pars compacta; SNr, Substantia nigra pars reticulata; VTA, ventral tegmental area. Scale bars: 200 μm (**a**); 5 μm (**b**); 100 μm (**c**); 10 μm (**d**); 50 μm (**e**).



Extended Data Figure 5 | Retrograde tracing with retrobeads.

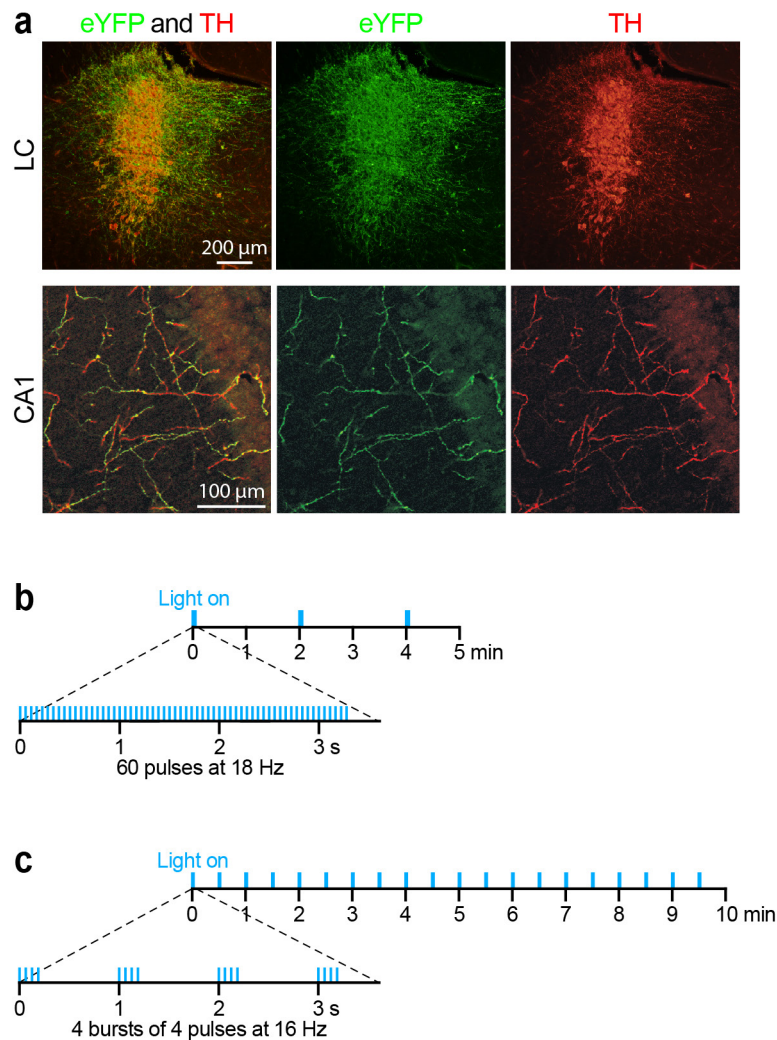
Representative images of coronal sections containing VTA (a, b) and LC (c, d) showing TH⁺ (green) neurons labelled with retrobeads (red) in LC, but not in VTA. CA1 and CA3, hippocampal subregions CA1 and CA3; DG, dentate gyrus; LC, locus coeruleus; MPB, medial parabrachial

nucleus; SNc, Substantia nigra pars compacta; SNr, Substantia nigra pars reticulata; SuM, supramammillary nucleus; VTA, ventral tegmental area. Scale bars: 200 μ m (a); 50 μ m (b–d). The mouse brain in this figure has been reproduced with permission from Franklin, K. B. J. & Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates* 3rd ed, 691 (Academic, 2007).



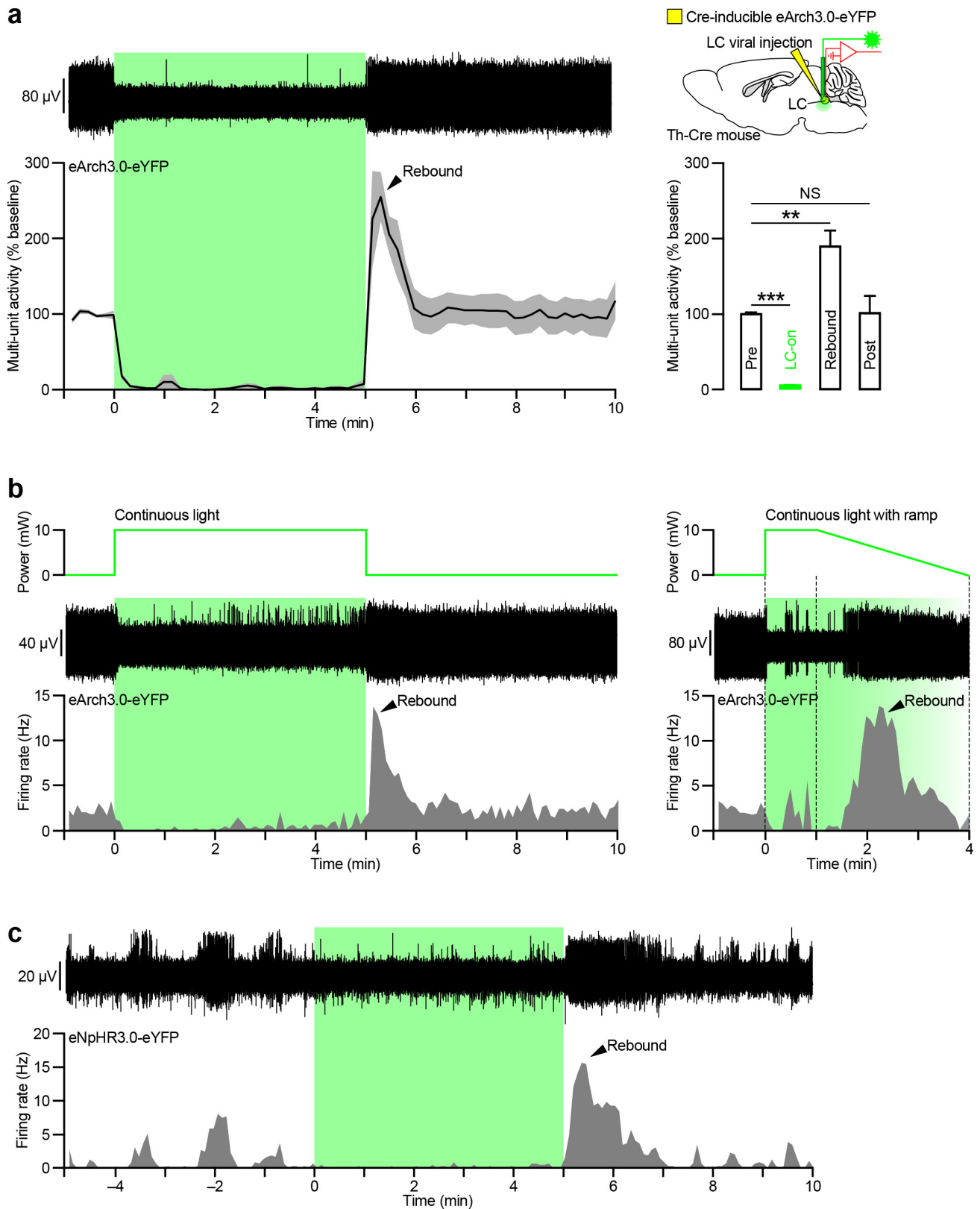
Extended Data Figure 6 | Performance during training and probe tests in the optogenetic activation experiment. **a**, ChR2-expressing LC-TH⁺ and VTA-TH⁺ neurons reliably follow 25-Hz blue light stimulation in awake mice. **b**, ChR2⁺ mice ($n = 8$) and ChR2⁻ controls ($n = 6$) both acquired the task over several weeks of training and maintained exceptionally stable performance from session (S) 46 until the end of training (S46 to S125: group \times session interaction, $F_{15,180} = 1.14$, $P > 0.05$; group effect, $F_{1,12} = 4.63$, $P > 0.05$). After 15 training sessions, Th-Cre mice were allocated into ChR2⁺ and ChR2⁻ groups based on their performance (average performance index in S1–S15: 75% in both groups). Vertical

grey bar denotes the break in training due to implantation surgery. Pre, pre-training. **c**, Ten-minute probe test conducted before the implantation surgery, showing good memory (correct digging: one-sample t -test versus chance, $t_{13} = 3.17$, $P < 0.01$; ChR2⁺ versus ChR2⁻: two-tailed paired t -test, $t_{12} < 1$, $P > 0.05$). **d**, Infusion of propranolol (Prop) or SCH23390 (SCH) in dorsal hippocampus 20 min before LC photoactivation had no impact on the total digging time in either group (ChR2⁺: $F_{3,21} < 1$, $P > 0.05$; ChR2⁻: $F_{3,15} = 1.44$, $P > 0.05$) (see Fig. 4f). ** $P < 0.01$ versus chance. Dashed lines indicate the chance level. Data are means \pm s.e.m.



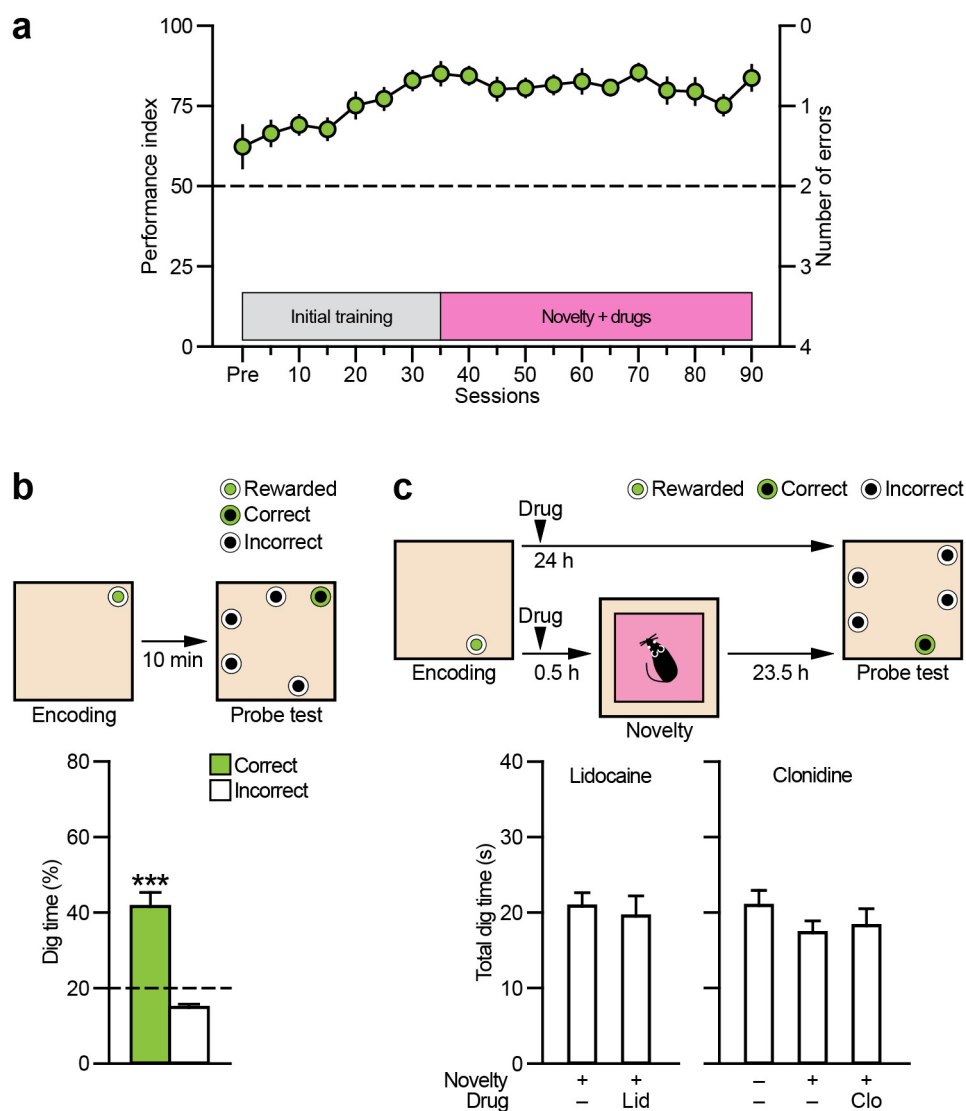
Extended Data Figure 7 | ChR2-eYFP expression in LC-TH⁺ neurons of Th-Cre mice and optogenetic stimulation protocols for *ex vivo* hippocampal electrophysiology. **a**, Representative images of double immunofluorescence for eYFP (green) and TH (red) in LC cell bodies

(top) and LC axons in CA1 (bottom) in the strain of Th-Cre mice used for *ex vivo* hippocampal electrophysiology. **b**, **c**, Photostimulation protocols used in intracellular (**b**) and extracellular (**c**) *ex vivo* hippocampal electrophysiology.



Extended Data Figure 8 | Rebound activation after optogenetic inhibition of LC-TH⁺ neurons. **a**, *In vivo* optrode recordings of eArch3.0-expressing LC-TH⁺ neurons in anaesthetized Th-Cre mice. Left, complete inhibition of multi-unit activity in LC during the 5 min '532-nm light on' period (example trace and population data ($n = 9$ traces from 5 mice)), followed by pronounced rebound activation. Grey shading represents \pm s.e.m. Right, mean multi-unit activity in LC (Pre versus LC-on: $t_8 = 35.6$, $P < 0.001$; Pre versus rebound: $t_8 = 3.98$, $P < 0.01$; Pre versus Post: $t_8 < 1$, $P > 0.05$). NS, not significant. ** $P < 0.01$, *** $P < 0.001$. **b**, Single traces

showing rebound excitation of eArch3.0-expressing LC-TH⁺ neurons after incomplete inhibition. Rebound excitation could not be reduced by low-power illumination (10 mW, left), even when light intensity was decreased monotonically over 3 min (right). **c**, Rebound excitation of the membrane trafficking-enhanced variants of halorhodopsin (eNpHR3.0)-expressing LC-TH⁺ neurons of Th-Cre mouse. Green shading indicates 'light on' periods. The mouse brain in this figure has been reproduced with permission from Franklin, K. B. J. & Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates* 3rd ed, 691 (Academic, 2007).



Extended Data Figure 9 | Performance during training and probe tests in the pharmacological inactivation experiment. **a**, Th-Cre mice ($n = 15$) acquired the task over several weeks of training and maintained stable performance from session (S) 36 until the end of training (S36–S90: $F_{10,140} < 1$, $P > 0.05$). Pre, pre-training. **b**, Mice showed effective memory when tested 1 h after weak encoding (one-sample t -test versus chance,

$t_{14} = 5.68$, $P < 0.001$). **c**, Intra-VTA infusion of lidocaine (Lid, left) or intraperitoneal injection of clonidine (Clo, right) before novelty exploration had no impact on the total digging time (see Fig. 6b, c) (lidocaine: $t_{14} < 1$, $P > 0.05$; clonidine: $F_{2,28} = 1.01$, $P > 0.05$). *** $P < 0.001$ versus chance. Dashed lines indicate the chance level. Data are means \pm s.e.m.

Activation mechanism of endothelin ET_B receptor by endothelin-1

Wataru Shihoya^{1,2,3*}, Tomohiro Nishizawa^{3,4*}, Akiko Okuta^{2*‡}, Kazutoshi Tani², Naoshi Dohmae⁵, Yoshinori Fujiyoshi^{1,2}, Osamu Nureki³ & Tomoko Doi⁶

Endothelin, a 21-amino-acid peptide, participates in various physiological processes, such as regulation of vascular tone, humoral homeostasis, neural crest cell development and neurotransmission. Endothelin and its G-protein-coupled receptor are involved in the development of various diseases, such as pulmonary arterial hypertension, and thus are important therapeutic targets. Here we report crystal structures of human endothelin type B receptor in the ligand-free form and in complex with the endogenous agonist endothelin-1. The structures and mutation analysis reveal the mechanism for the isopeptide selectivity between endothelin-1 and -3. Transmembrane helices 1, 2, 6 and 7 move and envelop the entire endothelin peptide, in a virtually irreversible manner. The agonist-induced conformational changes are propagated to the receptor core and the cytoplasmic G-protein coupling interface, and probably induce conformational flexibility in TM6. A comparison with the M2 muscarinic receptor suggests a shared mechanism for signal transduction in class A G-protein-coupled receptors.

Endothelin-1 (ET-1) was discovered in the supernatant of cultured aortic endothelial cells in 1988, and is the most potent, long-lasting vasoconstrictor ever discovered in humans^{1,2}. Intensive studies of ET-1 indicated that ET-1 is a widely distributed, multifunctional hormone that works in both paracrine and autocrine manners². ET-1 and its related isopeptides, ET-2 and ET-3, perform several physiological functions in neural crest development, cell proliferation, sodium excretion, salt homeostasis, and regulation of vascular tone and cell growth^{3–5}. They transmit signals through two receptor subtypes, the ET_A and ET_B receptors, which share approximately 60% sequence similarity^{4,6} and belong to the class A G-protein-coupled receptors (GPCRs).

Both receptors bind ET-1 in a ‘quasi-irreversible’ manner with sub-nanomolar affinities^{7–10}, but they transmit different signals via multiple G proteins and β -arrestin^{11–13}. In essence, the ET_A and ET_B receptors exert opposing actions on various physiological processes, such as vasoregulation and cell growth. For example, in the vascular system, the ET_A receptor performs the primary vasoconstricting role, and because of its irreversible binding with ET-1, it mediates extremely prolonged vasoconstriction^{3,5}. On the other hand, the ET_B receptor mainly induces NO-mediated vasorelaxation and functions as the ‘clearance receptor’ that scavenges the circulating ET-1 via the lysosomal pathway^{3,5,14}. The endothelin system, especially the ET_A receptor is involved in the development of diseases and pathological conditions, such as arterial hypertension, atherosclerosis, heart failure, renal disease, diabetes and cancer^{2–5,15}. Therefore, the ET_A receptor is a particularly important target for developing treatments for these diseases^{15,16}, and indeed non-selective/ET_A-selective antagonists bosentan, macitentan and ambrisentan have been clinically used for the treatment of pulmonary arterial hypertension^{17–19}. In addition, the ET_B-selective agonist IRL-1620 (*N*-succinyl-(Glu⁹, Ala^{11,15})-ET-1_{8–21}) is also in clinical trials to enhance delivery of anti-tumour drugs by improving blood flow to the tumour²⁰. However, despite numerous pharmacological and

biochemical studies on the ET receptors and the ET isopeptides, little is known about how these receptors bind their peptide agonists and transmit signals. Here we present the crystal structures of the ET_B receptor, in the complex with an endogenous agonist ET-1 and in the ligand-free form. These structures, together with the mutational analysis, reveal the molecular mechanism for agonist binding, isopeptide selectivity and signal transduction by ET_B, and provide insights into the general mechanism of signal transduction by the class A GPCRs.

Overall structures

For structural study, we employed a thermostabilized ET_B receptor containing five mutations (ET_BR-Y5: R124Y^{1,55}, D154A^{2,57}, K270A^{3,55}, S342A^{6,54}, and I381A^{7,48} (superscripts indicate Ballesteros–Weinstein numbers²¹))²² (Extended Data Fig. 1a, b and Supplementary Discussion ‘Thermostabilized ET_B receptor’). Using slightly different constructs, we successfully crystallized and determined the structures of the ET_B receptor in complex with the endogenous agonist ET-1 (ET-1-bound structure) and in the absence of ligand (ligand-free structure), at 2.8 and 2.5 Å resolutions, respectively (Fig. 1a, b, Extended Data Fig. 2a–f, Extended Data Table 1 and Methods). The ET_B receptor adopts the typical GPCR architecture, comprising seven transmembrane (TM) helices and intracellular helix 8 (H8). The extracellular loop (ECL)2 forms long anti-parallel β -strands with a short hairpin, which is a common feature of the peptide receptors²³.

We first describe the ET-1-bound structure, to clarify the detailed interactions between ET-1 and the ET_B receptor. The extracellular portion of the receptor is widely involved in ET-1 binding. The amino (N)-terminal tail, the three extracellular loops (ECL1–ECL3), and the six TM helices (TM2–TM7) altogether constitute the orthosteric pocket, which is entirely occupied by ET-1 (Fig. 1a). The interacting surface area between ET-1 and the receptor encompasses up to $\sim 1,500$ Å², and is much more extensive than that of neurotensin

¹Department of Basic Medicinal Sciences, Graduate School of Pharmaceutical Sciences, Nagoya University, Chikusa, Nagoya 464-8601, Japan. ²Cellular and Structural Physiology Institute, Nagoya University, Chikusa, Nagoya 464-8601, Japan. ³Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, Bunkyo, Tokyo 113-0032, Japan. ⁴Precursory Research for Embryonic Science and Technology (PRESTO), Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan. ⁵Biomolecular Characterization Unit, RIKEN Center for Sustainable Resource Science, Wako, Saitama 351-0198, Japan. ⁶Department of Biophysics, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan.

*These authors contributed equally to this work.

‡Deceased.

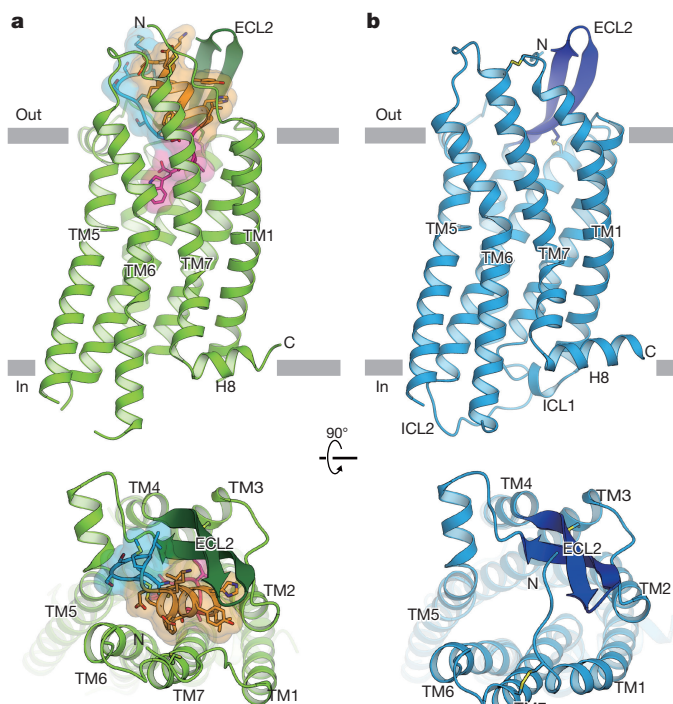


Figure 1 | ET-1 bound and ligand-free structures of ET_B receptor.

a, b, Overall structure of ET_B in complex with ET-1 (**a**, yellow-green) and in the ligand-free form (**b**, cyan), viewed from within the membrane plane (top panels) and from the extracellular side (bottom panels). ET_B is depicted by ribbons, with the ECL2 β -sheet highlighted in deep colours. The disulfide bonds at the N terminus and ECL2 of ET_B are shown in yellow sticks. ET-1 is shown as transparent surface representation and a ribbon model, with its N-terminal region in cyan, α -helical region in orange and C-terminal region in pink. The side chains of ET-1 are shown as sticks.

receptor-1 (NTSR1) bound to its truncated agonist peptide NTS₈₋₁₃ (ref. 24) (~ 800 Å²) and even broader than that of the chemokine receptor CXCR4 bound to the virus chemokine vMIP-II²⁵ ($\sim 1,300$ Å²) (Extended Data Fig. 3a–d). Such a wide-ranging interaction could account for the exceptionally high affinity for ET-1 with an apparent dissociation constant (K_d) on the sub-nanomolar order^{6,26}. ET-1 adopts a bicyclic structure with two intrachain disulfide bond pairs (C1–C15 and C3–C11), as observed in the monomer structures of the ET-1 peptide analysed by nuclear magnetic resonance (NMR) and X-ray crystallography^{27–29} (Fig. 2a and Extended Data Fig. 4a; amino-acid residues of ET-1 are denoted by one-letter codes throughout). The central residues, D8 to L17 (α -helical region), form approximately

three α -helical turns, and the C1 to M7 residues (N-terminal region) are tightly anchored to the α -helical region by the disulfide bond pairs. The D18 to W21 residues (carboxy (C)-terminal region) adopt extended conformations and deeply penetrate into the receptor core, with the C-terminal W21 side chain directed towards the depths of the pocket.

As suggested in the previous monomer structures of ET-1 and ET-like peptides, the N-terminal and α -helical regions of ET-1 adopt rather stable conformation (Extended Data Fig. 4a–e). In contrast, the C-terminal region, which is essential for the agonist potency of ET-1 (ref. 30), was flexible and almost disordered in the previous monomer structures^{28,29}. In the complex structure, these C-terminal ‘tail’ residues penetrate into the receptor core and are specifically recognized by the receptor through more than 16 residues (Extended Data Figs 5a, b and 6a–c). In addition, owing to the bicyclic architecture of the ET-1 peptide, the N-terminal amino group is in the close vicinity of the C-terminal carboxylate group and the D18 side chain (4.1 and 4.0 Å, respectively), where it contributes to the stability of the C-terminal region (Fig. 2a).

Binding interaction of ET-1

The three C-terminal residues of ET-1, I19, I20 and W21, fit into the hydrophobic pocket formed by the hydrophobic residues within the depths of the orthosteric pocket of ET_B (Ile157^{2,60}, Pro178^{3,29}, Val185^{3,36}, Phe240^{ECL2}, Leu277^{5,42}, Trp336^{6,48}, Leu339^{6,51} and Tyr369^{7,36}) (Fig. 2b). The backbone amide and carbonyl of I20 are hydrogen-bonded by Asn158^{2,61} and Gln181^{3,32}, respectively, for further stabilization. The C-terminal carboxylate and the D18 side chain form an electrostatic interaction network that involves the charged residues of ET_B (Lys182^{3,33}, Lys273^{5,38}, Arg343^{6,55} and D368^{7,35}). The C-terminal W21 side chain makes a π -cation interaction with the ϵ -amino group of Lys182^{3,33}, and concurrently directly interacts with Trp336^{6,48} in the CWXP motif (Fig. 2b), which is thought to be involved in the signalling function in class A GPCRs. These interactions are particularly important for the agonistic potency of ET-1, as the deletion or substitution of W21 results in the loss of receptor binding and activation³¹. The observed interactions of ET-1 explain the strict structural requirements of the C-terminal region for the binding ability and agonistic potency³².

The α -helical region primarily comprises hydrophobic residues and is sandwiched between the ECL2 and TM6–7 of ET_B, where it forms extensive van der Waals interactions with the receptor (Fig. 2c and Extended Data Figs 5a, b and 6b). F14 of ET-1 interacts with the hydrophobic residues on TM7 of ET_B (Leu361^{7,28}, Leu364^{7,31} and Leu365^{7,32}), while V12 and Y13 of ET-1 interact with the hydrophobic residues on ECL2 of ET_B (Ile243, Met245, Tyr247, Leu252, Ile254 and Leu256). The long extended β -sheet of ECL2 tightly holds ET-1,

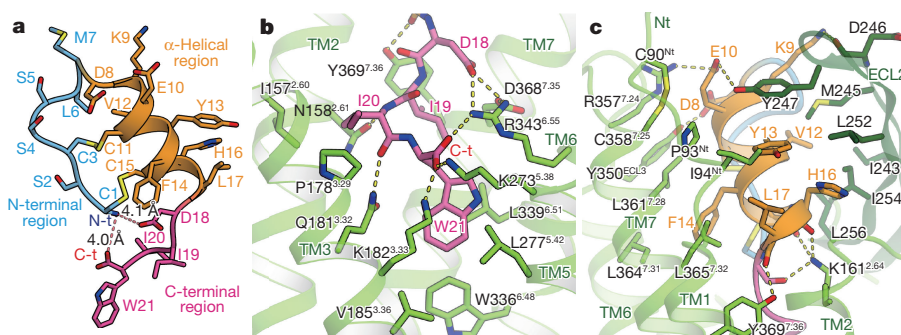


Figure 2 | Orthosteric pocket of ET_B receptor. **a**, The architecture of ET-1 bound to ET_B receptor is shown in a ribbon representation, with its side chains, N-terminal amino group (N-t) and C-terminal carboxyl group (C-t) shown as sticks. Each domain is coloured as in Fig. 1a. The N-terminal amino group is in close vicinity to the carboxyl groups of the

C terminus and D18 side chain. Distances between the amino and carboxyl groups are indicated. **b, c**, Detailed interactions between ET-1 and ET_B at the C-terminal region (**b**) and the α -helical region (**c**). Hydrogen bonds are shown as yellow dotted lines.

and at the tip of ECL2, Tyr247^{ECL2} caps the upper side of ET-1 by making a π - π stacking interaction with Y13 (Fig. 2c). In addition to the hydrophobic interactions, the charged residues of ET-1 at the N-terminal end of the α -helical region interact with ECL2 and ECL3 of ET_B (Fig. 2c): D8 and K9 form hydrogen bonds with the side chain of Tyr350^{ECL3} and the backbone carbonyl of Asp246^{ECL2}, respectively, and E10 interacts with the side chains of Tyr247^{ECL2} and Arg357^{7.24}. At the C-terminal end of the α -helical region, the backbone carbonyls of ET-1 are coordinated by ET_B residues: C15, H16 and D18 interact with Lys161^{2.64} and L17 interacts with Tyr369^{7.36} (Fig. 2c).

To verify the biological relevance of this complex structure, we investigated the binding of ET-1 to the mutant ET_B receptors, in which the residues for ET-1 binding were substituted with alanine. Consistent with the previous biochemical studies^{33–36}, mutations at residues involved in the interactions with the α -helical and C-terminal regions of ET-1 (Lys182^{3.33}, Lys273^{5.38}, Leu277^{5.42}, Leu339^{6.51}, Arg357^{7.24}, Asp368^{7.35} and Leu365^{7.32}) have significant effects on ET-1 binding, supporting the importance of their direct interactions with ET-1 (Extended Data Fig. 7a, b). However, the decreases in the affinities of these mutants were within the range of one order of magnitude, suggesting that ET-1 binding to the receptor is extensively supported by numerous well-arranged interactions. On the other hand, these mutant receptors revealed considerably larger defects in binding of the ET_B-specific isopeptide ET-3 (Extended Data Fig. 7c), suggesting that, while the tight associations between the ETs and the ET_B receptor are essentially mediated through the α -helical and C-terminal regions, the N-terminal region can affect and modulate these interactions (Supplementary Discussion ‘Isopeptide selectivity’).

ET-1-induced conformational changes

The structural comparison between the ligand-free and ET-1-bound ET_B receptors revealed dramatic agonist-induced rearrangements in the extracellular orthosteric pocket (Fig. 3a). Compared with the ligand-free structure, the extracellular portions of TM2, TM6 and TM7 move inwards by about 2.6, 4.1 and 4.9 Å, respectively, in the ET-1-bound structure. By contrast, TM1, which has no direct interaction with ET-1, moves outwards by about 4.4 Å. Associated with the movement of these helices, the orthosteric pocket contracts and adopts a compact ‘closed’ configuration in the ET-1-bound structure, which enables tight interactions with ET-1 (Fig. 3b, c). In addition, the N-terminal tail and the ECL2 β -sheet of the receptor together form a lid-like architecture that covers the orthosteric pocket (Figs. 2c and 3c). ET-1 forms an incredibly stable complex with both ET receptor subtypes, as its dissociation rate is of the order of days^{8,9}. The ‘closed’ configuration and the lid-like architecture of the ET-1-bound structure would account for the virtually irreversible binding of ET-1. Such an irreversibly binding manner is likely to be conserved in the ET_A receptor, where the agonist–receptor complex mediates extremely prolonged signalling unless affected by other molecules that allosterically induce ligand dissociation^{7,10,37}. In contrast, in the ligand-free structure, the surrounding TM helices are loosely packed against each other, thereby creating a large cavity that extends to the receptor core (Fig. 3d). The resultant broad entrance might facilitate the access of large peptide agonists to the orthosteric pocket.

The structural comparison provides further mechanistic insights into the ET-1-induced activation of the ET_B receptor. The ET-1 residues at positions 18 and 19 are closely related to its agonistic potency. For example, the (Thr¹⁸, γ -methyl-Leu¹⁹)-substituted ET-1 analogue functions as a non-selective antagonist for endothelin receptors, while retaining high affinity³⁸. Since D18 and I19 of ET-1 are involved in the interactions with TM6 and TM7, they are probably responsible for the inward motion of the extracellular portions of TM6 and TM7 (Fig. 4a). Therefore, the ET-1-induced inward shifts of these helices are likely to be the critical step for the activation of the ET_B receptor. The ligand–receptor interaction and consequent conformational changes in the

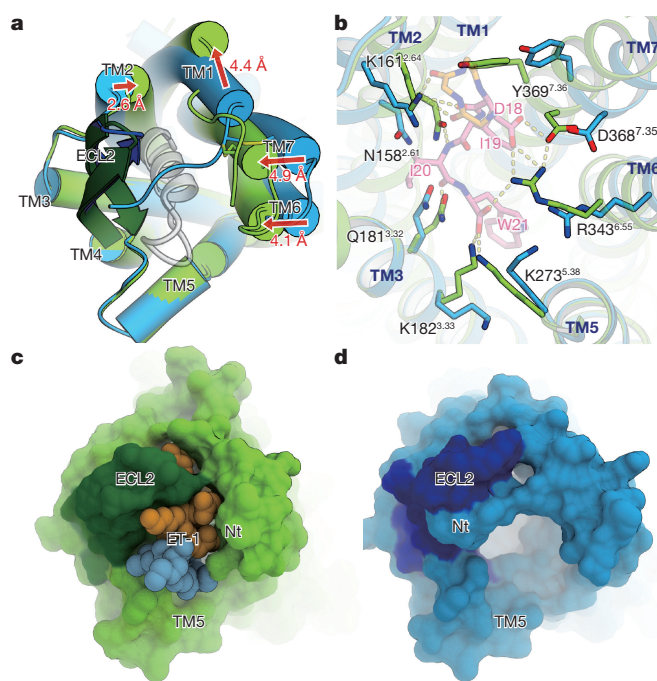


Figure 3 | Structural comparison between ET-1-bound and ligand-free ET_B receptors. **a**, Superimposition of the ET-1-bound and ligand-free structures of the ET_B receptor on residues 113–301 of the ET_B receptor. The extracellular view shows the ET-1-induced movement of the TM helices that constitute the orthosteric pocket. Red arrows indicate the movement of helices, with the distances of the residues at each end of the helix (TM1: Glu98; TM2: Leu162; TM6: Leu349; TM7: Cys358). **b**, Close-up view of the orthosteric pocket. The ET_B residues are attracted towards the centre of the receptor and form hydrogen-bonding and electrostatic interactions with ET-1. **c**, **d**, Surface representation of the ET-1-bound (**c**) and ligand-free (**d**) structures of ET_B viewed from the extracellular side. ET-1 is shown as a CPK model, coloured as in Fig. 1a.

ET_B receptor bound to ET-1 are reminiscent of those of the nanobody-stabilized active M2 acetylcholine receptor (M2R) bound to the agonist and the positive allosteric modulator (PAM)³⁹ (Fig. 4b), in which the allosteric coupling between the orthosteric pocket and the cytoplasmic G-protein coupling interface is mediated through the TM6 helix. One of the common features of the reported active GPCR structures is the outward displacement of the cytoplasmic segment of TM6, which creates the cytoplasmic cavity for G-protein binding (Extended Data Figs 8 and 9b–e). In M2R, this motion is likely to be coupled to the agonist-induced inward shift of the extracellular segment of TM6 (Fig. 4b). In the current ET-1-bound ET_B receptor, the extracellular portion of TM6 shifted inwards and is tightly stabilized through interactions with ET-1 (Fig. 4a). However, the cytoplasmic segment of TM6 is tilted towards the inside of the receptor and packed against the surrounding TM helices, where it probably adopts an inactive conformation (Extended Data Fig. 9a). In addition, the conserved E/DRY motif forms an intrahelical salt bridge between Asp198^{3.49} and Arg199^{3.50}, which is the structural hallmark of an inactive GPCR.

A closer inspection of the current structures of the ET_B receptor revealed the ET-1-induced conformational changes in the receptor core. In the ET-1-bound structure, owing to the inward shift of the extracellular portion of TM6, the side chain of Trp336^{6.48} moves downwards by about 2.5 Å, with its C α atom shifted inwards by about 1.5 Å (Fig. 4c). Its position is further stabilized by the direct interaction with W21 of ET-1, resulting in the closer packing of Trp336^{6.48} against Phe332^{6.44}. Accordingly, the hydrophobic residues in the receptor core are tightly packed against each other in the ET-1-bound structure, compared with the ligand-free structure (Extended Data Fig. 10a–d). The hydrophobic contact interaction between Trp6.48 and Phe6.44

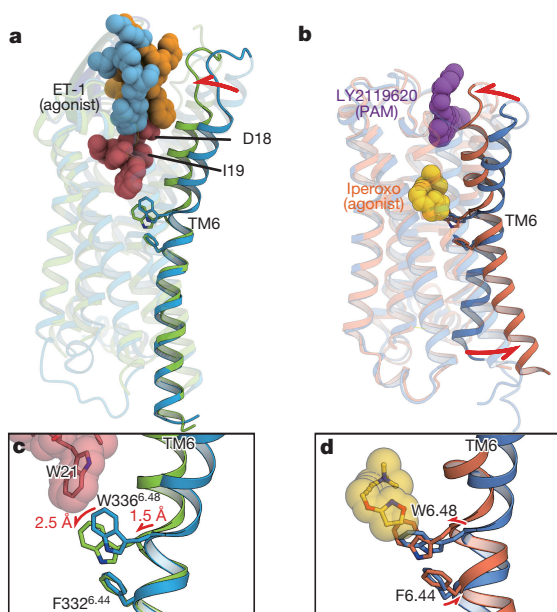


Figure 4 | Allosteric coupling through TM6. **a, b,** Conformational changes of TM6, shown in ribbon models. **a,** ET_B structures in the ligand-free and ET-1-bound forms are coloured cyan and yellow-green, respectively. ET-1 is represented as a CPK model. **b,** M2R structures bound to an antagonist (QNB) (PDB accession number 3UON) and an agonist (Iperoxo) and PAM (LY2119620) (PDB accession number 4MQT) are represented in blue and red, respectively. The agonist and PAM are shown in CPK models, while the antagonist is omitted for clarity. **c, d,** Close-up views of the conserved residues of TM6 are shown for ET_B receptor (**c**) and M2R (**d**). The Trp6.48 and Phe6.44 residues are represented by sticks. Red arrows indicate conformational changes upon agonist binding.

is implicated in the reorganization of the TM6 segments, and thus in the receptor activation, in NTSR1 (ref. 40) and μ -opioid receptor⁴¹. In addition, Trp6.48 commonly exhibits a slight downward shift upon activation in GPCRs with available active and inactive structures^{39,41,42} (Fig. 4d and Extended Data Fig. 9c–e), although the exact conformational changes are not identical. Together, these observations suggest that, as implicated in M2R, the ET-1-induced inward-shift of the extracellular segment of TM6 is likely to be coupled to the receptor activation, by inducing the outward displacement of Phe332^{6.44}. However, the allosteric coupling between the extracellular and cytoplasmic segments of TM6 is rather weak in the ET_B receptor, allowing the separate motion of each segment. According to the NMR and molecular dynamics studies of the β_2 adrenergic receptor and the μ -opioid receptor, agonist binding only induces structural heterogeneity, and G-protein binding is necessary to stabilize the fully active conformation^{43,44}. Likewise, in the ET_B receptor, the outward displacement of the cytoplasmic segment of TM6 seems to be strongly coupled to the G-protein binding, rather than to the ET-1 binding. In the current ET-1-bound structure, the crystal packing contacts might have captured the cytoplasmic segment of TM6 in its inactive conformation (Extended Data Fig. 2g and h).

Allosteric coupling through TM7

In contrast to TM6, the ET-1-induced conformational changes in TM1, TM2 and TM7 are propagated to the cytoplasmic side, as a slight inward shift of the cytoplasmic end of TM7 and H8 (Fig. 5a). TM1, TM2 and TM7 are bridged through a polar interaction network involving conserved hydrophilic residues (Asn119^{1.50}, Asp147^{2.50}, Asn378^{7.45} and Asn382^{7.49}) and water molecules within the transmembrane domain, where TM7 is kinked at the conserved Pro383^{7.50} residue (Fig. 5b). Upon ET-1 binding, this polar interaction network undergoes a large reorganization, associated with the inward shift of the extracellular portions of TM2 and TM7 (Fig. 5c). Asp147^{2.50} moves downwards by about 2 Å and participates in the interaction with Asn119^{1.50} and Asn382^{7.49},

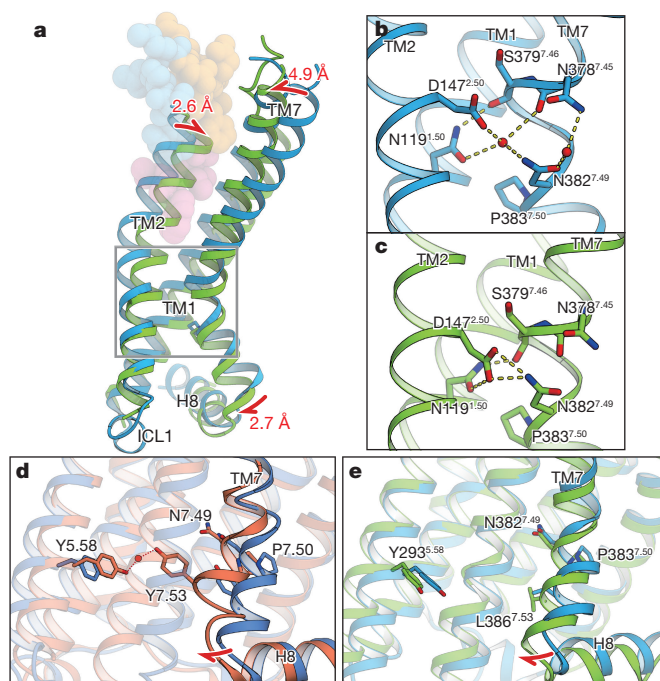


Figure 5 | Polar interaction network reorganization. **a–c,** Polar interaction network among TM1, TM2 and TM7. Structural changes of TM1, TM2, TM7 and H8 upon ET-1 binding are shown (**a**). Red arrows indicate the movement of each helix. Distances of the movement are measured for the C α position of the critical residue in each helix (TM2: L162^{2.65}; TM7: C358^{7.23}; TM8: S390^{7.56}). Close-up views of the polar interaction network among these TM helices are shown for ligand-free (**b**) and ET-1-bound (**c**) ET_B receptors. Water molecules are shown as red circles, and hydrogen bonding interactions are indicated by yellow dotted lines. **d, e,** TM7 and the NPXXY motifs in M2R (**d**) and ET_B receptor (**e**). Red arrows indicate movements of TM7 and H8. The Tyr7.53 residue is not conserved in ET_B. The putative water molecule is shown in the M2R structure.

where it replaces the hydrated water molecule that maintains the kinked conformation of TM7. Consequently, TM7 adopts a rather straight conformation in the ET-1 bound structure, with its cytoplasmic end shifted inwards by about 2.7 Å.

ET-1 binding causes another important change in the Na⁺ binding pocket that is conserved among class A GPCRs (Extended Data Fig. 10e and f). The negatively charged Asp147^{2.50} residue, together with Ser184^{3.35}, Thr188^{3.39} and Ser379^{7.46}, constitutes the putative Na⁺ binding pocket in the ligand-free structure at the bottom of the orthosteric pocket, just beside the polar interaction network. This Na⁺ pocket completely collapses upon ET-1 binding, in association with the inward shift of TM2. The Asp147^{2.50} residue is no longer accessible from the extracellular side in the ET-1-bound structure, and the space for Na⁺ is filled with the bulky side chain of His150^{2.53}. The collapse of the Na⁺ pocket is a common event that occurs upon receptor activation, and thus implicated in G-protein signalling⁴⁵. Therefore, the reorganization of the polar interaction network and the consequent inward shift of the cytoplasmic end of TM7 in the ET_B receptor are likely to be associated with the receptor activation.

Similar inward shifts of TM7 were reported in the active conformation structures of other class A GPCRs, including M2R³⁹, μ -opioid receptor⁴¹, β_2 adrenergic receptor⁴² and rhodopsin⁴⁶, in which Tyr7.53 in the NPXXY motif forms a water-mediated hydrogen bond with Tyr5.58 to stabilizes the active conformation (Fig. 5d and Extended Data Fig. 9c–e). However, the ET_B receptor lacks the Tyr7.53 residue in the NPXXY motif, where it is replaced by leucine, and therefore the equivalent hydrogen bond cannot be formed even in the active conformation (Fig. 5e). Nonetheless, the inward shift of TM7 is likely

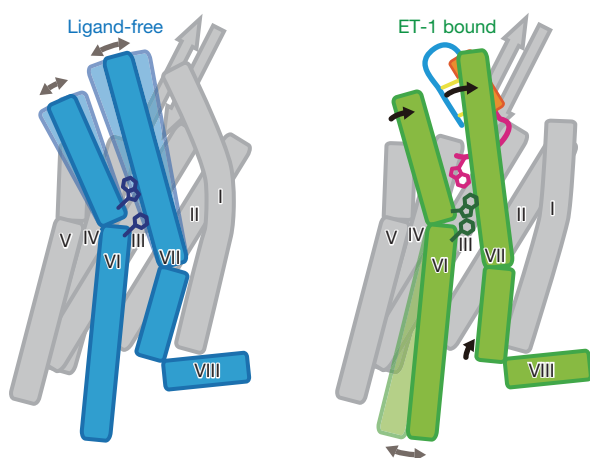


Figure 6 | Schematic representation of ET_B receptor activation by ET-1. Schematic representation of ET_B receptor in the ligand-free (left) and ET-1-bound (right) states. TM6, TM7 and H8 are highlighted and coloured according to the current crystal structures in Fig. 1a, b. Hydrophobic residues involved in the signal transduction (Trp6.48 and Phe6.44) are represented with sticks. Black arrows indicate the conformational changes in TM6 and TM7 upon ET-1 binding. The extracellular portions of TM6 and TM7 are flexible in the ligand-free state. ET-1 binding stabilizes the extracellular portions of these helices and evokes conformational changes in the receptor core and the cytoplasmic end of TM7, which consequently induce the outward displacement of the cytoplasmic segment of TM6.

to be conserved in the ET_B receptor. This slight shift of TM7 affects interactions between TM6 and TM7 (Extended Data Fig. 10g and h). Therefore, as well as the extracellular inward shift of TM6, the cytoplasmic inward shift of TM7 probably induces the mobility of the cytoplasmic segment of TM6, and thus should be a trigger that promotes receptor activation (Fig. 6). Overall, the current ET-1 bound ET_B receptor represents an agonist-bound, partly active state in which the ET-1-induced conformational changes are partly propagated to the cytoplasmic side and engage the receptor for subsequent G-protein binding.

It is also notable that ET-1 binding alone evokes the structural change in TM7, without evoking the outward shift of the cytoplasmic segment of TM6, in a similar manner to the agonist-bound A_{2A} receptor structure⁴⁷ (Extended Data Fig. 9f). NMR studies of the μ -opioid receptor revealed that the allosteric coupling between the orthosteric pocket and the G-protein-coupling interface preferentially occurred in TM1, ICL1, TM2 and TM7, rather than in TM6 (ref. 44). Therefore, this manner of signal transduction might be somewhat conserved among the class A GPCRs, while their agonists and detailed conformational changes are quite diverse. These findings will advance our understanding of the molecular mechanisms of signal transduction in peptide-activated and other class A GPCRs.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 December 2015; accepted 20 July 2016.

Published online 5 September 2016.

- Yanagisawa, M. *et al.* A novel potent vasoconstrictor peptide produced by vascular endothelial cells. *Nature* **332**, 411–415 (1988).
- Barton, M. & Yanagisawa, M. Endothelin: 20 years from discovery to therapy. *Can. J. Physiol. Pharmacol.* **86**, 485–498 (2008).
- Kedzierski, R. M. & Yanagisawa, M. Endothelin system: the double-edged sword in health and disease. *Annu. Rev. Pharmacol. Toxicol.* **41**, 851–876 (2001).
- Kohan, D. E., Rossi, N. F., Inscho, E. W. & Pollock, D. M. Regulation of blood pressure and salt homeostasis by endothelin. *Physiol. Rev.* **91**, 1–77 (2011).
- Rubanyi, G. M. & Polokoff, M. A. Endothelins: molecular biology, biochemistry, pharmacology, physiology, and pathophysiology. *Pharmacol. Rev.* **46**, 325–415 (1994).

- Davenport, A. P. International Union of Pharmacology. XXIX. Update on endothelin receptor nomenclature. *Pharmacol. Rev.* **54**, 219–226 (2002).
- Desmarests, J., Gresser, O., Guedin, D. & Frelin, C. Interaction of endothelin-1 with cloned bovine ETA receptors: biochemical parameters and functional consequences. *Biochemistry* **35**, 14868–14875 (1996).
- Mey, J. G. R. D., Compeer, M. G. & Meens, M. J. Endothelin-1, an endogenous irreversible agonist in search of an allosteric inhibitor. *Mol. Cell. Pharmacol.* **1**, 246–257 (2009).
- Takasaka, T., Sakurai, T., Goto, K., Furuichi, Y. & Watanabe, T. Human endothelin receptor ET_B. Amino acid sequence requirements for super stable complex formation with its ligand. *J. Biol. Chem.* **269**, 7509–7513 (1994).
- Hilal-Dandan, R., Villegas, S., Gonzalez, A. & Brunton, L. L. The quasi-irreversible nature of endothelin binding and G protein-linked signaling in cardiac myocytes. *J. Pharmacol. Exp. Ther.* **281**, 267–273 (1997).
- Doi, T., Sugimoto, H., Arimoto, I., Hiroaki, Y. & Fujiyoshi, Y. Interactions of endothelin receptor subtypes A and B with G_i, G_o, and G_q in reconstituted phospholipid vesicles. *Biochemistry* **38**, 3090–3099 (1999).
- Maguire, J. J. *et al.* Comparison of human ET_A and ET_B receptor signalling via G-protein and β -arrestin pathways. *Life Sci.* **91**, 544–549 (2012).
- Rosano, L. *et al.* β -arrestin links endothelin A receptor to β -catenin signaling to induce ovarian cancer cell invasion and metastasis. *Proc. Natl Acad. Sci. USA* **106**, 2806–2811 (2009).
- Bremnes, T. *et al.* Regulation and intracellular trafficking pathways of the endothelin receptors. *J. Biol. Chem.* **275**, 17596–17604 (2000).
- Rosano, L., Spinella, F. & Bagnato, A. Endothelin 1 in cancer: biological implications and therapeutic opportunities. *Nature Rev. Cancer* **13**, 637–651 (2013).
- Remuzzi, G., Perico, N. & Benigni, A. New therapeutics that antagonize endothelin: promises and frustrations. *Nature Rev. Drug Discov.* **1**, 986–1001 (2002).
- Clozel, M. *et al.* Pathophysiological role of endothelin revealed by the first orally active endothelin receptor antagonist. *Nature* **365**, 759–761 (1993).
- Sidharta, P. N., van Giersbergen, P. L., Halabi, A. & Dingemans, J. Macitentan: entry-into-humans study with a new endothelin receptor antagonist. *Eur. J. Clin. Pharmacol.* **67**, 977–984 (2011).
- Vatter, H. & Seifert, V. Ambrisentan, a non-peptide endothelin receptor antagonist. *Cardiovasc. Drug Rev.* **24**, 63–76 (2006).
- Maguire, J. J. & Davenport, A. P. Endothelin@25 - new agonists, antagonists, inhibitors and emerging research frontiers: IUPHAR Review 12. *Br. J. Pharmacol.* **171**, 5555–5572 (2014).
- Ballesteros, J. A. Integrated methods for the construction of three dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci.* **25**, 366–428 (1995).
- Okuta, A., Tani, K., Nishimura, S., Fujiyoshi, Y. & Doi, T. Thermostabilization of the human endothelin type-B receptor. *J. Mol. Biol.* **428**, 2265–2274 (2016).
- Venkatakrishnan, A. J. *et al.* Molecular signatures of G-protein-coupled receptors. *Nature* **494**, 185–194 (2013).
- White, J. F. *et al.* Structure of the agonist-bound neurotensin receptor. *Nature* **490**, 508–513 (2012).
- Qin, L. *et al.* Structural biology. Crystal structure of the chemokine receptor CXCR4 in complex with a viral chemokine. *Science* **347**, 1117–1122 (2015).
- Chen, J., Sawyer, N. & Regan, L. Protein-protein interactions: general trends in the relationship between binding affinity and interfacial buried surface area. *Protein Sci.* **22**, 510–515 (2013).
- Janes, R. W., Peapus, D. H. & Wallace, B. A. The crystal structure of human endothelin. *Nature Struct. Biol.* **1**, 311–319 (1994).
- Takahima, H. *et al.* Distributed computing and NMR constraint-based high-resolution structure determination: applied for bioactive Peptide endothelin-1 to determine C-terminal folding. *J. Am. Chem. Soc.* **126**, 4504–4505 (2004).
- Andersen, N. H., Chen, C. P., Marschner, T. M., Krystek, S. R. Jr & Bassolino, D. A. Conformational isomerism of endothelin in acidic aqueous media: a quantitative NOESY analysis. *Biochemistry* **31**, 1280–1295 (1992).
- Lättig, J., Oksche, A., Beyermann, M., Rosenthal, W. & Krause, G. Structural determinants for selective recognition of peptide ligands for endothelin receptor subtypes ET_A and ET_B. *J. Pept. Sci.* **15**, 479–491 (2009).
- Tam, J. P. *et al.* Alanine scan of endothelin: importance of aromatic residues. *Peptides* **15**, 703–708 (1994).
- Galantino, M. *et al.* D-amino acid scan of endothelin: importance of amino acids adjacent to cysteinyl residues in isomeric selectivity. *Pept. Res.* **8**, 154–159 (1995).
- Saeki, T., Ihara, M., Fukuroda, T., Yamagiwa, M. & Yano, M. [Ala^{1,3,11,15}] endothelin-1 analogs with ET_B agonistic activity. *Biochem. Biophys. Res. Commun.* **179**, 286–292 (1991).
- Nakajima, K. *et al.* Structure-activity relationship of endothelin: importance of charged groups. *Biochem. Biophys. Res. Commun.* **163**, 424–429 (1989).
- Ergul, A., Tackett, R. L. & Puett, D. Identification of receptor binding and activation sites in endothelin-1 by use of site-directed mutagenesis. *Circ. Res.* **77**, 1087–1094 (1995).
- Lee, J. A. *et al.* Lysine 182 of endothelin B receptor modulates agonist selectivity and antagonist affinity: evidence for the overlap of peptide and non-peptide ligand binding sites. *Biochemistry* **33**, 14543–14549 (1994).
- Talbot, A. *et al.* Aspirin and sodium salicylate inhibit endothelin ETA receptors by an allosteric type of mechanism. *Mol. Pharmacol.* **57**, 797–804 (2000).

38. Kikuchi, T. *et al.* Endothelin-1 analogues substituted at both position 18 and 19: highly potent endothelin antagonists with no selectivity for either receptor subtype ET_A or ET_B. *J. Med. Chem.* **36**, 4087–4093 (1993).
 39. Kruse, A. C. *et al.* Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature* **504**, 101–106 (2013).
 40. Krumm, B. E., White, J. F., Shah, P. & Grisshammer, R. Structural prerequisites for G-protein activation by the neurotensin receptor. *Nature Commun.* **6**, 7895 (2015).
 41. Huang, W. *et al.* Structural insights into μ -opioid receptor activation. *Nature* **524**, 315–321 (2015).
 42. Rasmussen, S. G. *et al.* Crystal structure of the β_2 adrenergic receptor-Gs protein complex. *Nature* **477**, 549–555 (2011).
 43. Nygaard, R. *et al.* The dynamic process of β_2 -adrenergic receptor activation. *Cell* **152**, 532–542 (2013).
 44. Sounier, R. *et al.* Propagation of conformational changes during μ -opioid receptor activation. *Nature* **524**, 375–378 (2015).
 45. Katritch, V. *et al.* Allosteric sodium in class A GPCR signaling. *Trends Biochem. Sci.* **39**, 233–244 (2014).
 46. Standfuss, J. *et al.* The structural basis of agonist-induced activation in constitutively active rhodopsin. *Nature* **471**, 656–660 (2011).
 47. Xu, F. *et al.* Structure of an agonist-bound human A_{2A} adenosine receptor. *Science* **332**, 322–327 (2011).
- 22227004, 24227004, 25650019, 26440024, and 26640102; the Core Research for Evolutional Science and Technology Program; Platform for Drug Discovery, Information, and Structural Life Science from the Ministry of Education, Culture, Sports, Science, and Technology of Japan; the Japan New Energy and Industrial Technology Development Organization (NEDO), the Japan Agency for Medical Research and Development (AMED) and the National Institute of Biomedical Innovation.
- Author Contributions** W.S. optimized the construct for crystallization, performed the thermostability assay, developed the purification procedure, expressed and purified the ET-1-bound and ligand-free receptors, and crystallized and collected data for the ligand-free structure. T.N. crystallized the ET-1-bound receptor and collected data, and solved and refined the structures. A.O. performed the thermostabilization screening and established the ET_BR-Y5 construct, and made the initial T4L-fused construct for crystallization. K.T. designed the T4L-fused construct and assisted with the structural determination. N.D. analysed the purified protein by mass spectrometry. T.D. performed the pharmacological analyses. The manuscript was prepared by W.S., T.N., K.T., Y.F., O.N. and T.D.; Y.F., O.N. and T.D. supervised the research.
- Author Information** Coordinates and structure factors have been deposited in the Protein Data Bank under accession number 5GLH for the ET-1-bound and 5GLI for the ligand-free ET_B receptors. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to O.N. (nureki@bs.s.u-tokyo.ac.jp) or T.D. (doi@mb.biophys.kyoto-u.ac.jp).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Nishimura and K. Kumazaki for discussions, Y. Oomae and K. Yamaguchi for ET_B mutant preparations, the beamline staff at BL32XU of SPring-8 (Hyogo, Japan) for technical help during data collection, and T. Suzuki for the mass spectrometry. This work was supported by Japan Society for the Promotion of Science KAKENHI grant numbers 15J09780,

Reviewer Information *Nature* thanks M. Barton, A. Davenport and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Thermostability assay of ET_BR-Y5 in detergent solution. The human ET_B receptor gene encoding a hexa-histidine tag in the amino-terminal tail (6hNET_BR)^{11,48} was used as the starting template, and we systematically surveyed stabilizing mutations according to the previously described method⁴⁹. The established thermostabilized ET_B receptor construct contained five mutations (ET_BR-Y5: R124Y^{1,55}, D154A^{2,57}, K270A^{5,35}, S342A^{6,54}, I381A^{7,48}). The thermostability of the ET_BR-Y5 construct was confirmed by a fluorescence detection size-exclusion chromatography-based thermostability assay (FSEC-TS)⁵⁰ as follows. The ET_BR-Y5 protein fused with green fluorescent protein (GFP) at the C terminus was expressed in SF⁺ insect cells (*expresSF⁺* cell, Protein Sciences). The cells were solubilized in a buffer containing 10 mM HEPES, pH 7.5, 200 mM NaCl and 1% *n*-dodecyl-β-D-maltopyranoside (DDM). After removing insoluble materials by centrifugation at 100,000g for 20 min, 100-μl aliquots of the supernatant were placed into polymerase chain reaction tubes and incubated at the respective temperatures for 10 min. The heat-treated samples were centrifuged at 100,000g for 20 min and loaded onto a Superdex200 10/150 column pre-equilibrated with 10 mM HEPES, pH 7.5, 200 mM NaCl and 0.05% DDM. Each fluorescent signal intensity at the monomeric peak was normalized to that of the unheated sample as 100% (Extended Data Fig. 1c). Each measurement was performed three times. Melting temperatures (*T_m*) were determined by fitting the curves to a sigmoidal dose-response equation, using the SigmaPlot 12 software (Systat Software).

Expression and purification of ET_B receptor in complex with ET-1. The haemagglutinin signal peptide, followed by the Flag epitope tag (DYKDDDD) and a nine-amino-acid linker, was added to the N terminus of the receptor, and a tobacco etch virus (TEV) protease recognition sequence was introduced between Gly57 and Leu66, to remove the disordered N terminus during the purification process. The C terminus was truncated after Ser407, and three cysteine residues were mutated to alanine (C396A, C400A and C405A) to avoid heterogeneous palmitoylation⁵¹. To improve crystallogenesis, T4 lysozyme containing the C54T and C97A mutations⁵² was introduced into intracellular loop 3, between Lys303^{5,68} and Leu311^{6,23} (ET_BR-Y5-T4L) (Extended Data Figs 1 and 2b).

The recombinant baculovirus was prepared using the Bac-to-Bac baculovirus expression system (Invitrogen). SF⁺ insect cells were infected with the virus at a cell density of 1.5×10^6 cells per millilitre in Sf900 II medium (Invitrogen), supplemented with 50 units ml⁻¹ penicillin, 50 μg ml⁻¹ streptomycin and 0.125 μg ml⁻¹ amphotericin B, and grown for 48 h at 27°C. The harvested cells were disrupted using a high-pressure homogenizer, EmulsiFlex-C5 (Avestin), in buffer containing 10 mM HEPES-NaOH, pH 7.5, 10 mM EDTA, 20% glycerol, 1 mM phenylmethylsulfonyl fluoride, 10 μg ml⁻¹ aprotinin, 10 μg ml⁻¹ leupeptin and 10 μg ml⁻¹ soybean trypsin inhibitor. The cell debris was removed by centrifugation at 4,00g for 30 min, and the crude membrane fraction was collected by ultracentrifugation at 100,000g for 1 h. The membrane fraction was solubilized in buffer, containing 50 mM HEPES-NaOH, pH 7.5, 200 mM NaCl, 2% DDM (Anatrace), 0.4% cholesterol hemisuccinate (CHS, Sigma) and 1 μM ET-1 (Peptide Institute Inc.), for 3 h at 4°C. Afterwards, 2 mg ml⁻¹ iodoacetamide was added to block reactive cysteines. The supernatant containing the solubilized receptor was separated from the insoluble material by ultracentrifugation at 100,000g for 1 h, and incubated with anti-Flag M2 affinity resin (Sigma) overnight at 4°C. After binding, the resin was washed with ten column volumes of wash I buffer, containing 20 mM HEPES-NaOH, pH 7.5, 500 mM NaCl, 0.1% lauryl maltose neopentyl glycol (LMNG, Anatrace) and 0.01% CHS, followed by ten column volumes of wash II buffer, containing 10 mM HEPES-NaOH, pH 7.5, 200 mM NaCl, 0.01% LMNG and 0.001% CHS. The receptor was eluted from the resin with 200 μg ml⁻¹ Flag peptide (Sigma) in the presence of 100 nM ET-1, and then treated with TEV protease (prepared in our laboratory) overnight at 4°C, to cleave the N-terminal flexible region of the receptor. The receptor was concentrated and loaded onto a Superdex200 10/300 size-exclusion column, equilibrated in buffer containing 10 mM HEPES-NaOH, pH 7.5, 200 mM NaCl, 0.003% LMNG, 0.0003% CHS and 100 nM ET-1. Peak fractions were pooled, concentrated to 40 mg ml⁻¹ using a Vivaspinn sample concentrator with a 50 kDa molecular mass cut-off (Sartorius), and frozen until crystallization. During the concentration, ET-1 was added to a final concentration of 100 μM.

Expression and purification of ligand-free ET_B receptor. The crystallization construct was further modified to obtain the ligand-free structure of ET_B (Extended Data Fig. 2a). T4L was modified according to the previous report⁵³, as follows. Amino-acid residues 13–60 were removed, and the linker sequence (–GGSGG–) was inserted at the corresponding site (ET_BR-Y5-mT4L). The EGFP–His₁₀ tag and the TEV protease cleavage site were introduced at the C terminus. The recombinant

baculovirus was prepared as described above. Sf9 insect cells were infected with the virus at a cell density of 4.0×10^6 cells per millilitre in Sf900 II medium, and grown for 72 h at 27°C. The harvested cells were disrupted by sonication, in buffer containing 20 mM Tris-HCl, pH 7.5, and 20% glycerol. The crude membrane fraction was collected by ultracentrifugation at 180,000g for 1 h. The membrane fraction was solubilized in buffer, containing 50 mM Tris-HCl, pH 7.5, 200 mM NaCl, 1% DDM, 0.2% cholesterol hemisuccinate and 2 mg ml⁻¹ iodoacetamide, for 2 h at 4°C. The supernatant was separated from the insoluble material by ultracentrifugation at 180,000g for 30 min, and incubated with TALON resin (Clontech) for 30 min. The resin was washed with ten column volumes of buffer, containing 20 mM Tris-HCl, pH 7.5, 500 mM NaCl, 0.1% LMNG, 0.01% CHS and 20 mM imidazole. The receptor was eluted in buffer, containing 20 mM Tris-HCl, pH 7.5, 200 mM NaCl, 0.1% LMNG, 0.01% CHS and 200 mM imidazole. The eluate was treated with TEV protease and dialysed against buffer (50 mM Tris-HCl, pH 7.5, and 500 mM NaCl). The cleaved GFP–His₁₀ tag and the TEV protease were removed with Ni²⁺-NTA resin. The receptor was concentrated and loaded onto a Superdex200 10/300 size-exclusion column, equilibrated in buffer containing 10 mM HEPES-NaOH, pH 7.5, 200 mM NaCl, 0.01% LMNG and 0.001% CHS. Peak fractions were pooled, concentrated to 40 mg ml⁻¹ using a centrifugal filter device (Millipore 50 kDa MW cutoff) and frozen until crystallization.

Crystallization. The purified receptors were reconstituted into the lipidic cubic phase (LCP) of monoolein (Nucheck), supplemented with cholesterol at a ratio of 4:5:1 (w/w) for protein:monoolein:cholesterol⁵⁴. The protein-laden mesophase was dispensed into 96-well glass plates in 30–50 nl drops and overlaid with 800–1,000 nl precipitant solution, using a mosquito LCP (TTP LabTech). Crystals of ET_BR-Y5-T4L in complex with ET-1 were grown at 20°C in the precipitant conditions containing 30% PEG400, 100 mM MES-NaOH (pH 6.0), 100 mM (NH₄)₂SO₄ and 5% 1,4-butanediol. The crystals of ET_BR-Y5-mT4L in the ligand-free form were grown in the precipitant conditions containing 18–25% PEG350MME, 100 mM MES-NaOH (pH 6.3), 80 mM (NH₄)₂SO₄, and 8% 1,4-butanediol. The crystals were harvested directly from the LCP using micromounts (MiTeGen) or LithoLoops (Protein Wave) and frozen in liquid nitrogen, without adding any extra cryoprotectant.

Data collection and structure determination. X-ray diffraction data were collected at the Spring-8 beamline BL32XU, using a 10 μm × 15 μm (width × height) micro-focused beam. Diffraction data were processed using XDS⁵⁵. The ET-1 bound structure was determined by molecular replacement with PHASER⁵⁶, using the T4L from chemokine receptor CXCR4 (PDB accession number 3OE0) and the poly-alanine model derived from the β₂-adrenergic receptor (PDB accession number 3NYA). Subsequently, the model was rebuilt and refined using COOT⁵⁷ and PHENIX⁵⁸, respectively. The ligand-free structure was determined by molecular replacement, using the ET-1 bound structure, and subsequently rebuilt and refined as described above. The final model of ET-1-bound ET_BR-Y5-T4L contained residues 88–129, 135–206, 217–303 and 311–401 of ET_B, all residues of T4L, all residues of ET-1, and 4 water molecules, and the model of ET_BR-Y5-mT4L contained residues 85–304 and 311–402 of ET_B, all residues of mT4L, 3 monoolein molecules, 4 sulfate ions and 24 water molecules. The model quality was assessed by MolProbity⁵⁹. Figures were prepared using cuemol (<http://www.cuemol.org/ja/>).

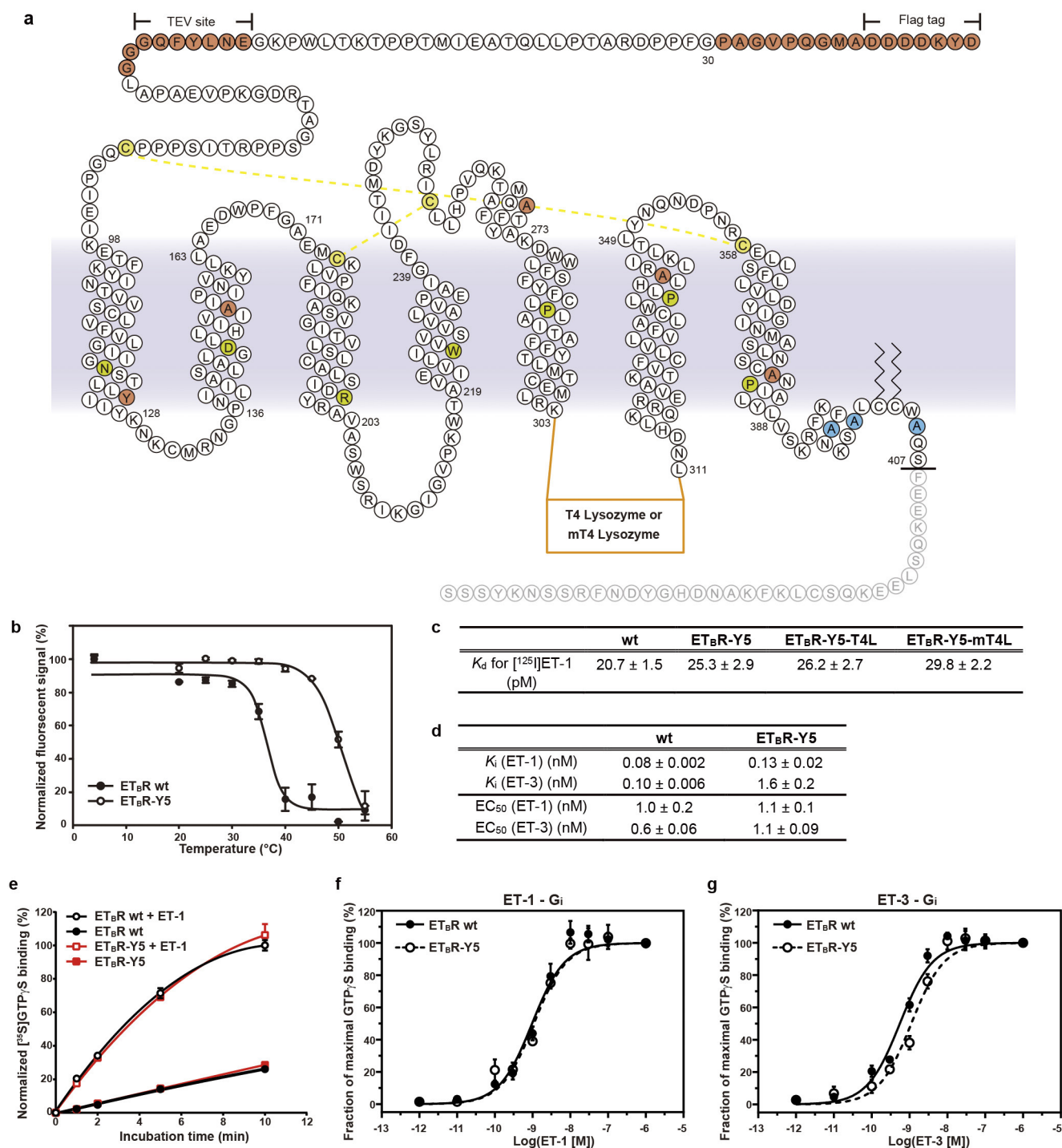
Radioligand binding studies of wild-type and mutant ET_B receptors. For competitive ligand binding assays, the genes encoding wild-type (6hNET_BR)¹¹ and mutant receptors were cloned in the pFastBac1 vector and the pCDNA3.1 vector, which were used for expression in insect cells and mammalian cells, respectively. Membranes from SF⁺ or HEK293 cells expressing 6hNET_BR or its mutants were prepared, and the expressed ET_B receptors were quantitated as described previously^{60,61}. Peptide binding competition was initiated by the addition of the membranes from SF⁺ cells (0.1–1.2 μg) or HEK293 cells (1–5 μg) to the assay mixture, composed of 0.1% bovine serum albumin (BSA), 0.03–0.05 nM [¹²⁵I]-labelled ET-1 (2,200 Ci mmol⁻¹, PerkinElmer Life Sciences), and eight concentrations of unlabelled ET-1 or ET-3 (ranging from 1 pM to 1 μM, Peptide Institute) in 50 mM HEPES-NaOH, pH 7.5, and 10 mM MgCl₂ (Mg-HEPES)⁶². Binding reactions were incubated at 37°C for 1 h, terminated by dilution with ice-cold Mg-HEPES, and filtered onto glass fibre filters in 96-well plates (multiscreen HTS FB, Merck Millipore) to separate the unbound [¹²⁵I]-labelled ET-1. After three washes with ice-cold Mg-HEPES, the radioactivity captured by the filters was counted using a γ counter. Filters were pretreated with 0.1% BSA in Mg-HEPES. The results were analysed by nonlinear regression, using the GraphPad Prism 6 software. In the saturation binding assays, membranes containing approximately 2 fmol receptor were incubated with six different concentrations of [¹²⁵I]-labelled ET-1, ranging from 1.5 pM to 253 pM in 100 μl of Mg-HEPES buffer containing 0.1% BSA, at 37°C for 2 h (ref. 62). The membranes were isolated from the unbound [¹²⁵I]-labelled ET-1 and washed, and the amount of receptor-bound [¹²⁵I]-labelled ET-1 was measured as described above. The non-specific binding of

the ^{125}I -labelled ET-1 in each reaction was assessed by including 100 nM ET-1 in the same reaction. The apparent dissociation constants (K_d) of ET-1 for wild-type ET_B receptor, ET_BR-Y5, and ET_BR-Y5-T4L were determined by fitting to a one-site binding equation, using the GraphPad Prism 6 software. Each experiment was performed 3 or 4 times.

G-protein activation assay. For the G protein activation assay, the ET_B receptor construct containing the haemagglutinin signal peptide, followed by the Flag sequence and hexa-histidine tag at the N terminus, was used as the wild type. Reconstitution of the purified receptor and G_i into phospholipid vesicles was performed as described previously¹¹. GDP/[^{35}S]GTP- γ S exchange assays were also performed as described previously, with 1.8 nM receptor, 50 nM G α_{i1} (purified from *E. coli*), ~140 nM G $\beta_1\gamma_2$ (purified from Sf9 cells), 1 μM GDP and 55 nM [^{35}S]GTP- γ S, with or without 1 μM ET-1, at 30 °C for the indicated times, in 20 mM HEPES-NaOH, pH 8.0, 1 mM EDTA, 100 mM NaCl, 10 mM MgCl₂ and 1 mM dithiothreitol. [^{35}S]GTP- γ S (1,250 Ci mmol⁻¹, PerkinElmer Life Sciences) was used after dilution with unlabelled GTP- γ S to 113.6 Ci mmol⁻¹. The reactions were terminated by addition with ice-cold stopping buffer containing 100 μM GTP, in 20 mM Tris-HCl, pH 8.0, 25 mM MgCl₂ and 100 mM NaCl, and filtered onto cellulose-mixed ester filters in 96-well plates (multiscreen HTS haemagglutinin, Merck Millipore), to isolate the G proteins from the unbound [^{35}S]GTP- γ S. After three washes with ice-cold stopping buffer without GTP, the radioactivity of the bound [^{35}S]GTP- γ S was measured, using a liquid scintillation counter. The measured values were normalized to that of wild-type receptor-activated [^{35}S]GTP- γ S binding for 10 min in the presence of ET-1. The assays were repeated twice. In the ligand concentration-dependent GDP/[^{35}S]GTP- γ S exchange assays, the receptors, reconstituted at 1 nM in the phospholipid vesicles with G proteins as described above, were pre-incubated in the 20 μl mixture with 11 different concentrations of ET-1 (ranging from 1 pM to 1 μM) for 5 min at 30 °C. The reactions were then started by the addition of [^{35}S]GTP- γ S. After an incubation for 2 min at 30 °C, the reactions were stopped by the addition of ice-cold stopping buffer, filtered and measured as described above. The assays were repeated four or five times. The data were analysed using the GraphPad Prism 6 software (Extended Data Fig. 1f, g).

48. Takasuka, T., Adachi, M., Miyamoto, C., Furuichi, Y. & Watanabe, T. Characterization of endothelin receptors ET_A and ET_B expressed in COS cells. *J. Biochem.* **112**, 396–400 (1992).

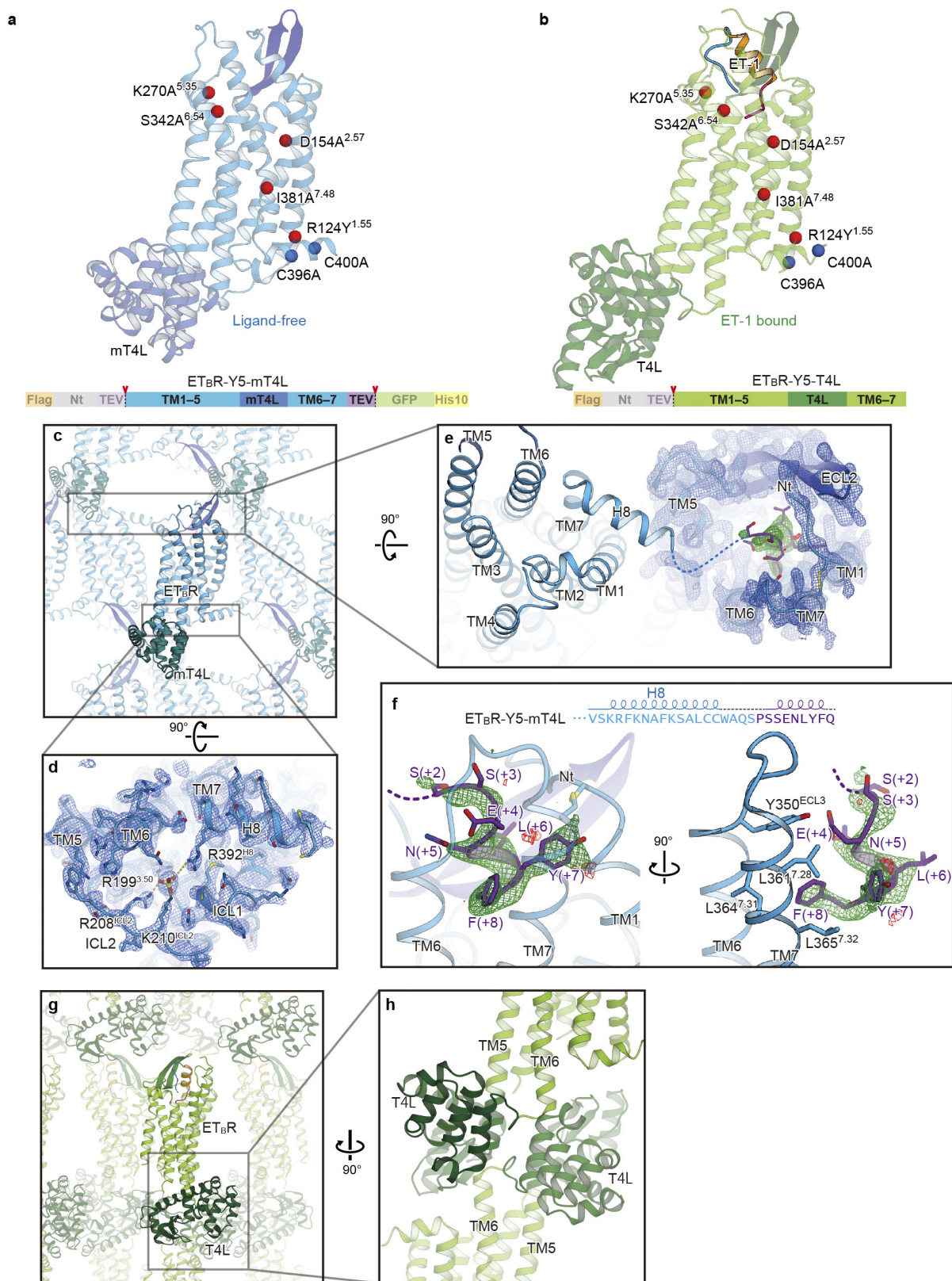
49. Serrano-Vega, M. J., Magnani, F., Shibata, Y. & Tate, C. G. Conformational thermostabilization of the β_1 -adrenergic receptor in a detergent-resistant form. *Proc. Natl Acad. Sci. USA* **105**, 877–882 (2008).
50. Hattori, M., Hibbs, R. E. & Gouaux, E. A fluorescence-detection size-exclusion chromatography-based thermostability assay for membrane protein precrystallization screening. *Structure* **20**, 1293–1299 (2012).
51. Okamoto, Y. *et al.* Palmitoylation of human endothelin. Its critical role in G protein coupling and a differential requirement for the cytoplasmic tail by G protein subtypes. *J. Biol. Chem.* **272**, 21589–21596 (1997).
52. Cherezov, V. *et al.* High-resolution crystal structure of an engineered human β_2 -adrenergic G protein-coupled receptor. *Science* **318**, 1258–1265 (2007).
53. Thorsen, T. S., Matt, R., Weis, W. I. & Kobilka, B. K. Modified T4 lysozyme fusion proteins facilitate G protein-coupled receptor crystallogenesis. *Structure* **22**, 1657–1664 (2014).
54. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nature Protocols* **4**, 706–731 (2009).
55. Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
56. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
57. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
58. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
59. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
60. Doi, T. *et al.* Characterization of human endothelin B receptor and mutant receptors expressed in insect cells. *Eur. J. Biochem.* **248**, 139–148 (1997).
61. Wada, K. *et al.* Purification of an endothelin receptor from human placenta. *Biochem. Biophys. Res. Commun.* **167**, 251–257 (1990).
62. Elshourbagy, N. A. *et al.* Molecular cloning and characterization of the major endothelin receptor subtype in porcine cerebellum. *Mol. Pharmacol.* **41**, 465–473 (1992).
63. Aumelas, A. *et al.* [Lys(-2)-Arg(-1)]endothelin-1 solution structure by two-dimensional ^1H -NMR: possible involvement of electrostatic interactions in native disulfide bridge formation and in biological activity decrease. *Biochemistry* **34**, 4546–4561 (1995).
64. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127–134 (1995).
65. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).



Extended Data Figure 1 | Thermostabilized construct of ET_B receptor.

a, Crystallization construct of ET_B receptor, shown with all of the modifications to the human wild-type ET_B receptor. The thermostabilizing mutations R124Y^{1,55}, D154A^{2,57}, K270A^{5,35}, S342A^{6,54} and I381A^{7,48}, and the three cysteine mutations, C396A, C400A and C405A, are coloured red and cyan, respectively. The C-terminal residues after S407 were truncated, and the T4L or mT4L was inserted between Lys303 and Leu311. The most conserved residues in each TM helix are coloured gold. Dashed lines indicate disulfide bonds. A Flag epitope tag was added after the N-terminal signal sequence, and a TEV protease site was introduced between Gly57 and Leu66. **b**, Thermostability profiles of the GFP-fused wild-type ET_B and ET_BR-Y5, measured by the FSEC-TS method⁵⁰. Each fluorescent signal intensity at the monomeric peak was normalized to that of the unheated sample as 100%. Data are given as means ± s.e.m. of three independent experiments. The wild-type ET_B-GFP (closed circles) has a melting temperature (T_m) of 36.7 °C and ET_BR-Y5-GFP (open circles)

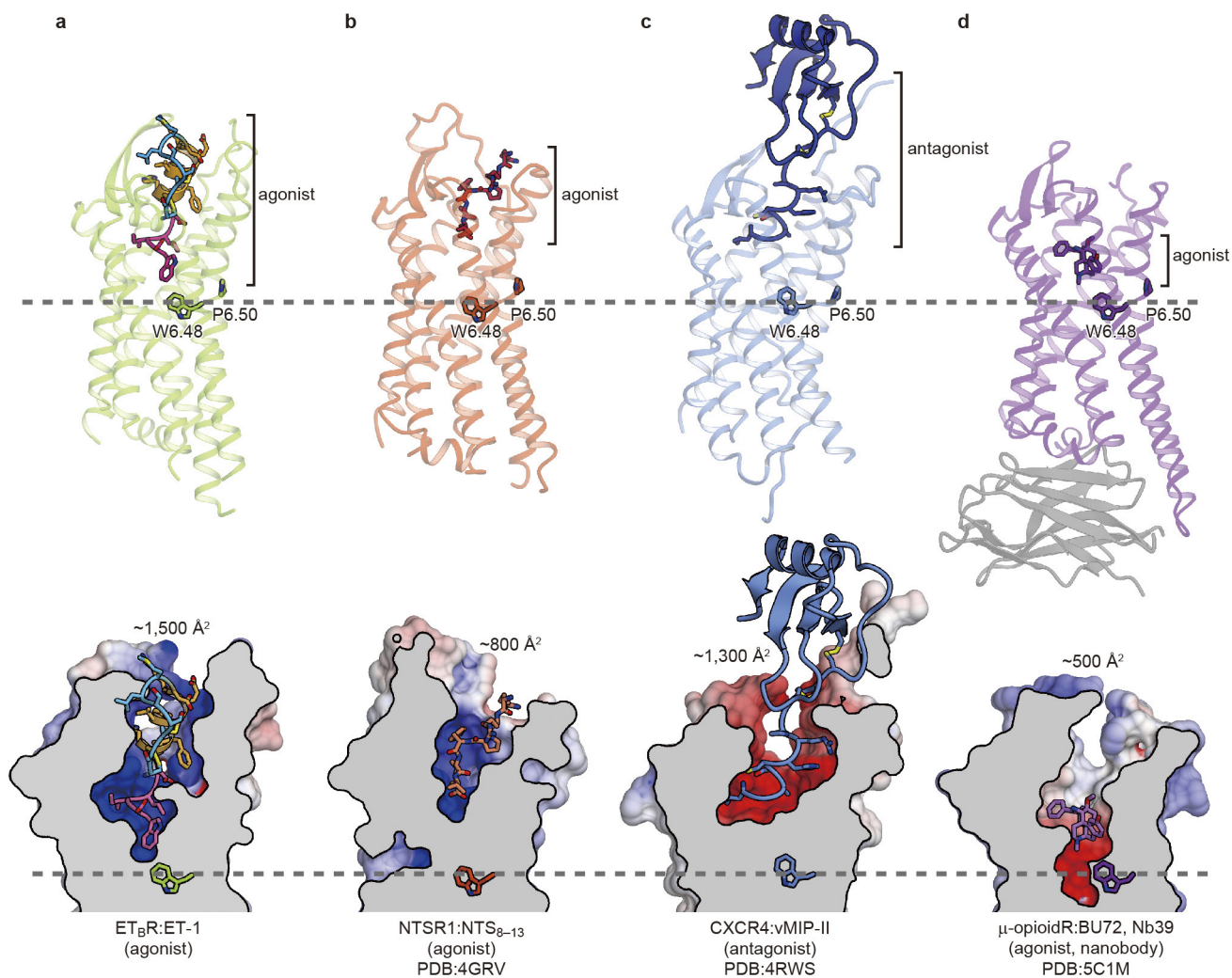
has a T_m of 50.1 °C, as calculated from the fitting curves. **c**, **d**, Apparent [¹²⁵I]-labelled ET-1 equilibrium dissociation constants (K_d). Values of the apparent dissociation constants for the wild-type (WT), thermostabilized (ET_BR-Y5), T4-fused (ET_B-Y4-T4L) and mT4L-fused (ET_BR-Y5-mT4L) constructs are shown. Each experiment was performed three or four times. **d**, The apparent inhibition constants (K_i) for [¹²⁵I]-labelled ET-1 binding and half-maximal effective concentrations (EC_{50}) for G_i activation by ET-1 and ET-3. Values for wild-type (WT) and thermostabilized (ET_BR-Y5) constructs are shown. **e**, Time courses for GTP-γS binding to the G protein G_i mediated by wild-type (circles) and thermostabilized (squares) ET_B receptors reconstituted into phospholipid vesicles, in the presence (open symbols) or absence (filled symbols) of ET-1. The assays were repeated four or five times. **f**, **g**, ET-1-dependent (**f**) and ET-3-dependent (**g**) G_i activation mediated by wild-type (closed circles) and thermostabilized (open circles) ET_B receptors.



Extended Data Figure 2 | See next page for caption.

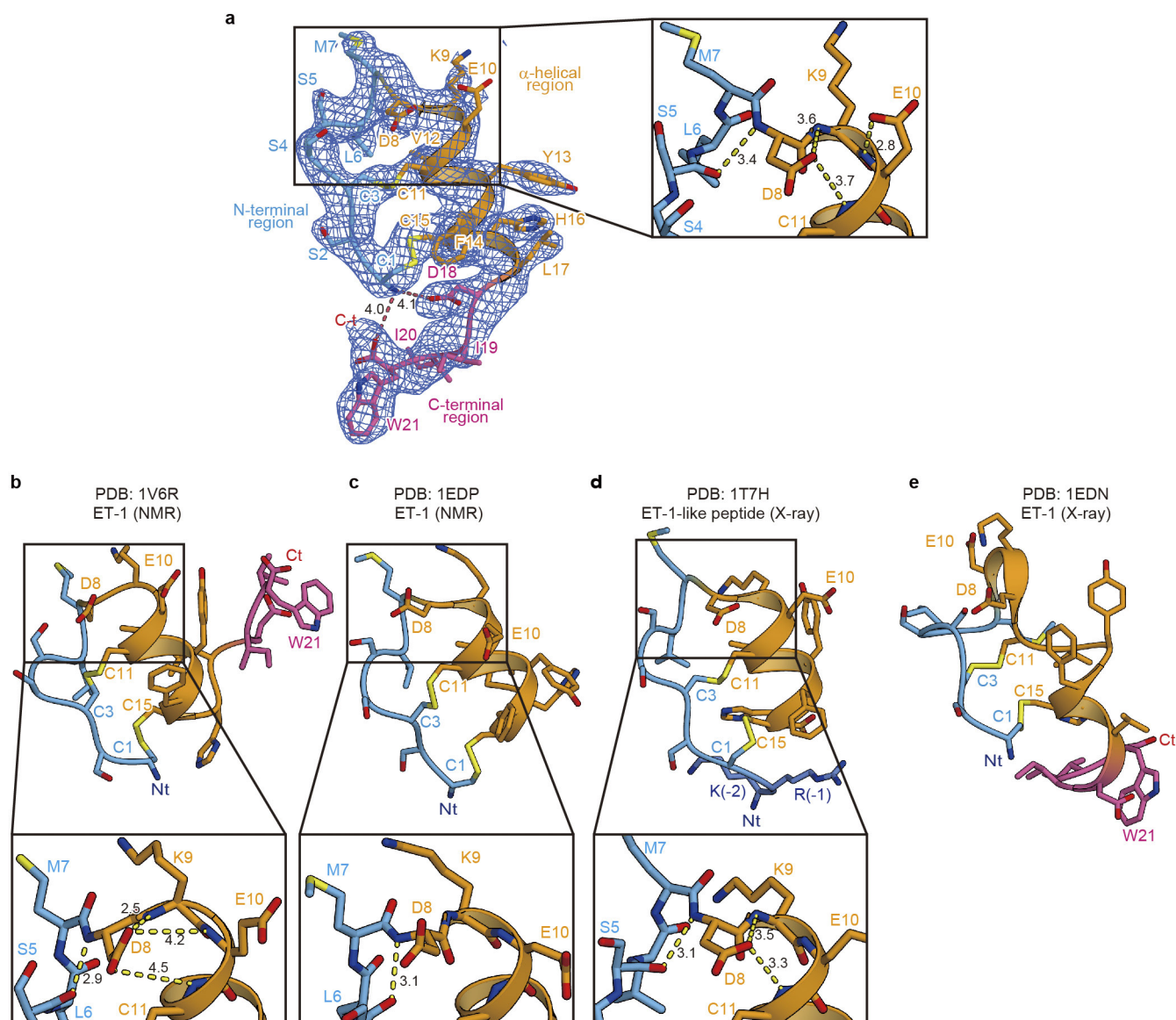
Extended Data Figure 2 | Ligand-free and ET-1-bound crystal structures of the ET_B receptor. **a, b**, Ligand-free (**a**) and ET-1-bound (**b**) structures of ET_B, and the crystallized constructs. Two crystal structures were obtained, using the different constructs indicated in each panel. The thermostabilizing mutations and the Cys-to-Ala mutations to avoid lipid modification are indicated with red and blue circles, respectively. **c–e**, Crystal packing of the ligand-free structure of ET_BR-Y5-mT4L (**c**) and the architectures of the cytoplasmic (**d**) and extracellular (**e**) sides. The $2F_o - F_c$ electron density map contoured at 0.8σ (blue mesh) revealed a sulfate ion bound to the cytoplasmic surface, which stabilizes the cytoplasmic architecture. The extracellular view shows a strong positive

$F_o - F_c$ density (green and red meshes contoured at 2.5 and -2.5σ , respectively) within the orthosteric pocket, which was assigned as the C-terminal tag residues from the adjacent molecule in the crystal lattice. **f**, Close-up extracellular view and the C-terminal tag sequence modelled in the density. These residues are not included in the deposited coordinate files, because we could not exclude the possibility that contaminant peptides are bound to the receptor. **g, h**, Crystal packing of the ET-1-bound structure of ET_BR-Y5-T4L (**g**) and close-up view of the crystal packing contacts between the adjacent molecules (**h**). TM6 is partly involved in the crystal packing with the adjacent molecule. TM5 forms a continuous helix, together with the first helix of T4L.



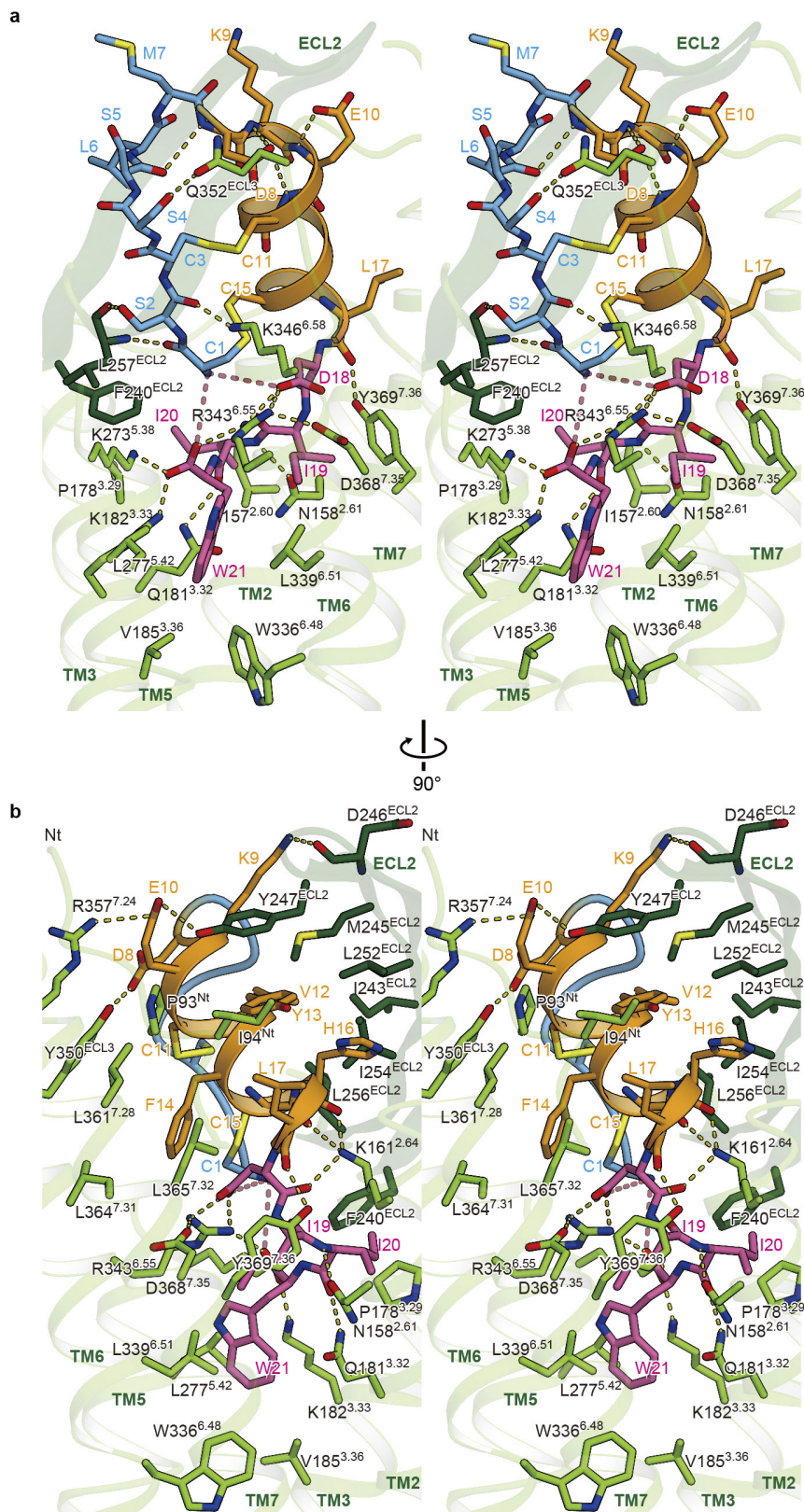
Extended Data Figure 3 | Structural comparison of the peptide-activated GPCRs. **a–d**, Comparison of the orthosteric pockets of the peptide-activated GPCRs bound to agonists (**a**, **b**, **d**) or an antagonist (**c**). Ribbon representations (top) and cutaway surfaces (bottom) for ET_B in complex with ET-1 (**a**), NTSR1 in complex with the NTS₈₋₁₃ peptide (PDB accession number 4GRV) (**b**), chemokine receptor CXCR4 in complex with the virus chemokine vMIP-II (PDB accession number 4RWS) (**c**) and the μ-opioid receptor in complex with the small-drug agonist BU72 and the nanobody Nb39 (PDB accession number 5C1M) (**d**) are aligned, according to the position of W6.48. The peptidic and small-drug agonist/

antagonist are represented by ribbons and sticks. Interaction ranges are indicated by black brackets (top), and the approximate interacting surface areas for their ligands are indicated (bottom). The extent of the penetration of the C-terminal tail of ET-1 is similar to the small drug-agonist (BU72) bound to the μ-opioid receptor, and is much deeper than the peptide agonist bound to NTSR1. The internal electric charges of the orthosteric pockets are complementary to the terminal charges of their peptide ligands: ET_B and NTSR1 are positively charged, while CXCR4 and the μ-opioid receptor are negatively charged.



Extended Data Figure 4 | Comparison of ET-1 structures. **a**, The $F_0 - F_c$ omit map for ET-1, contoured at 2.0σ , is shown. ET-1 is depicted by sticks and ribbons. The N-terminal end of the α -helical region is capped by the D8 and E10 side chains. The distances between the nitrogen at the N terminus and the carboxyl oxygens of D18 and W21 are indicated with red dotted lines (Å). The N-terminal and α -helical regions of ET-1 are stabilized by intra-peptide interactions; the negatively charged D8 and E10 side chains coordinate the backbone amides of the K9, E10 and C11 residues, supporting the α -helical folding, and the short hairpin at the M7 residue is stabilized by a hydrogen bond between the S6 carbonyl and the D8 amide. **b–e**, Reported structures of ET-1 and related peptides. NMR

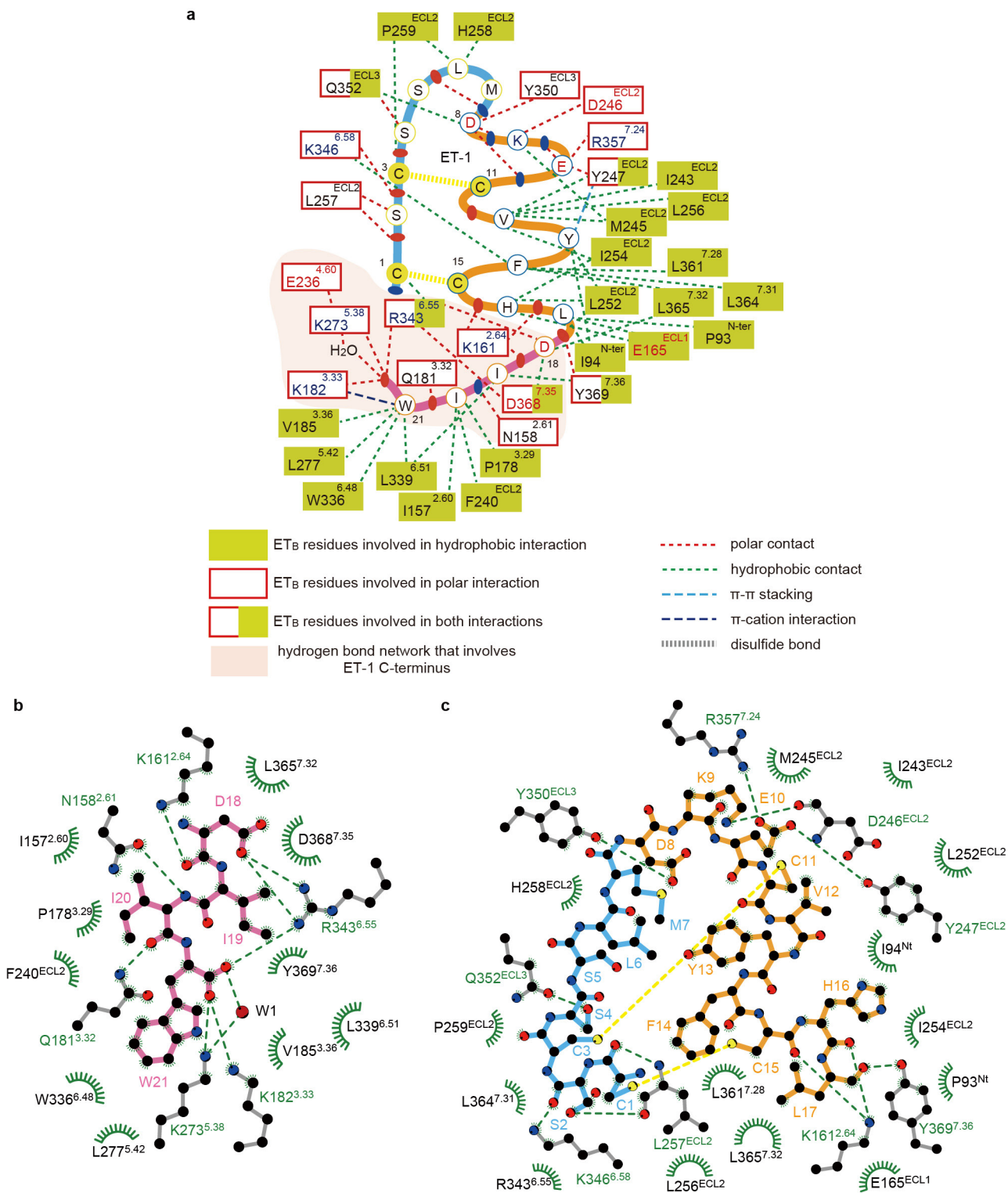
structures of ET-1: full-length model from 20 conformers²⁸ (**b**) (PDB accession number 1V6R) and a partial model that includes an unmodelled region (from L17 to W21)²⁹ (**c**) (PDB accession number 1EDP). X-ray crystal structure of the N-terminal-extended ET-1-like peptide⁶³ (PDB accession number 1T7H) (**d**). X-ray crystal structure of ET-1 (ref. 27) (PDB accession number 1EDN) (**e**). All structures are represented by sticks and ribbons, and the colour code is the same as in Fig. 1. The X-ray crystal structure of ET-1 (**e**) probably represents a rather deformed conformation affected by crystal packing interactions. Close-up views in **a–c** highlight the intra-peptide interactions that stabilize the common architecture of these peptides. Hydrogen bonds are indicated by yellow dotted lines with their respective distance values (Å).



Extended Data Figure 5 | Stereo views of the orthosteric pocket.

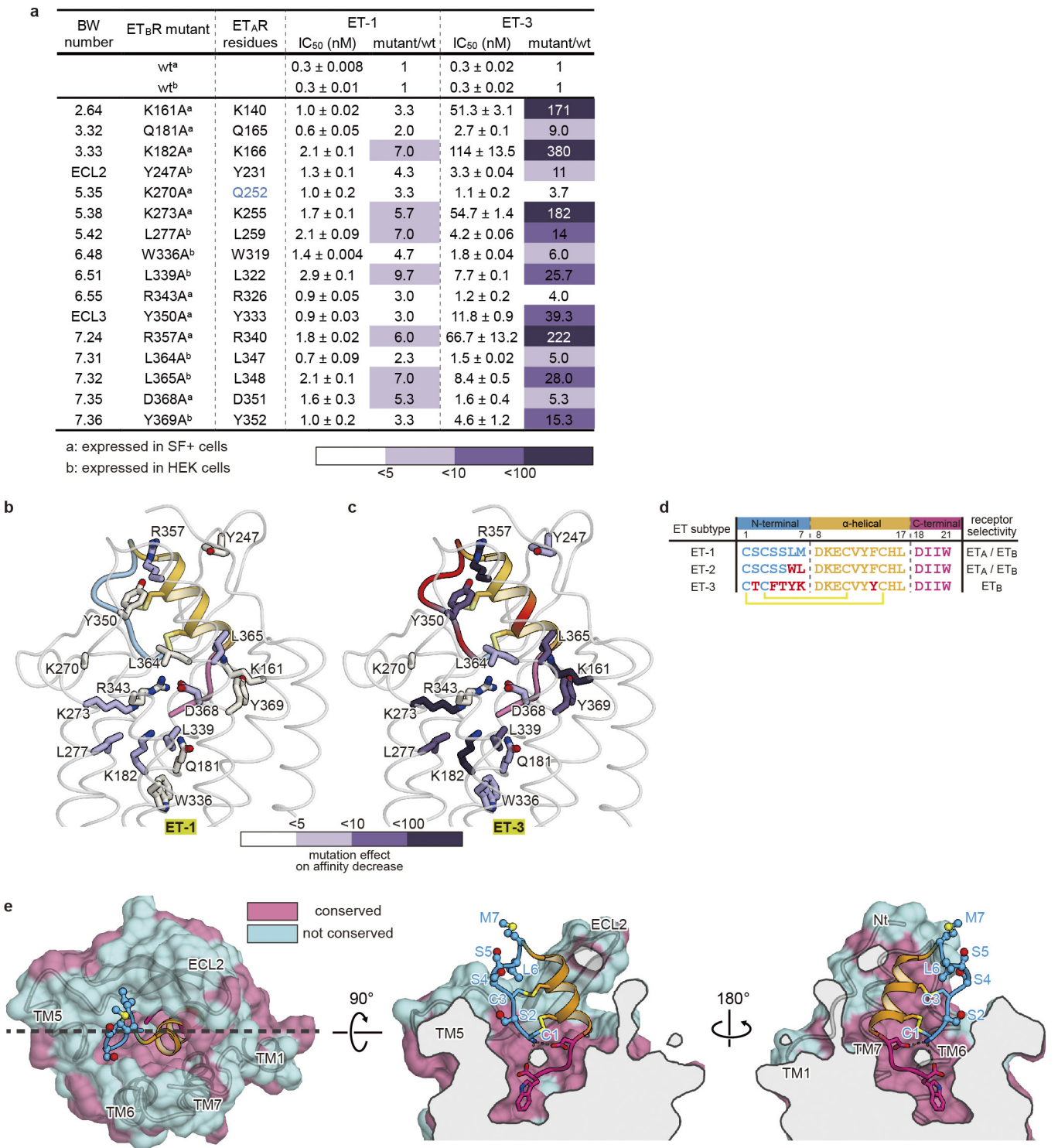
a, b, Stereo views showing the detailed interactions between ET-1 and ET_B receptor in the orthosteric pocket, viewed from different viewpoints. Residues involved in the major interactions between ET-1 and ET_B

receptor are shown, at the N-terminal and C-terminal regions of ET-1 (**a**) and at the α -helical and C-terminal regions (**b**). Hydrogen bonds are indicated with yellow dotted lines.



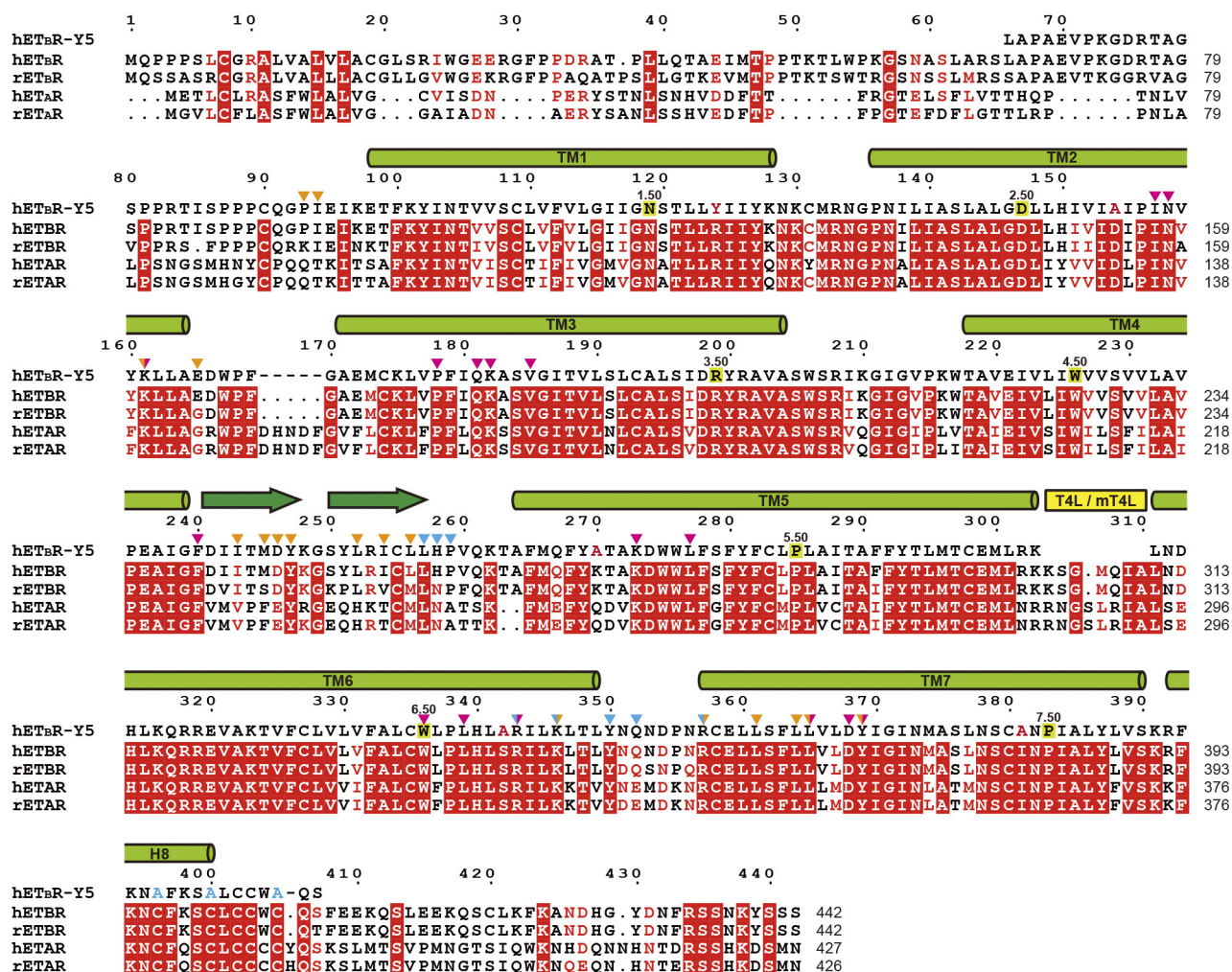
Extended Data Figure 6 | Ligand–receptor interactions. **a**, Schematic drawing of the orthosteric pocket. The residues shown here are within a radius of 4 Å around the ligand in the crystal structure. Amino-acid residues of ET-1 are represented by capital letters enclosed within circles. Blue and red ovals indicate main chain amide, and carbonyl and carboxyl groups of ET-1, respectively. All residues of the ET_B receptor involved in the interactions are indicated by large boxes and amino-acid letters, and the types of interaction are indicated with dotted lines. **b, c**, ET-1 interactions at the C-terminal region (**b**) and at the α-helical

and N-terminal regions (c) analysed by LIGPLOT⁶⁴. The labels and stick drawings of ET-1 residues are coloured cyan (N-terminal region), orange (α -helical region) and pink (C-terminal region), according to the same colour code used in Fig. 1. The ET_B receptor residues involved in the hydrophobic and hydrogen bond interactions are indicated by black and green letters, respectively. Intermolecular hydrogen bonds are indicated as green dashed lines, and disulfide bonds are indicated as yellow dashed lines.



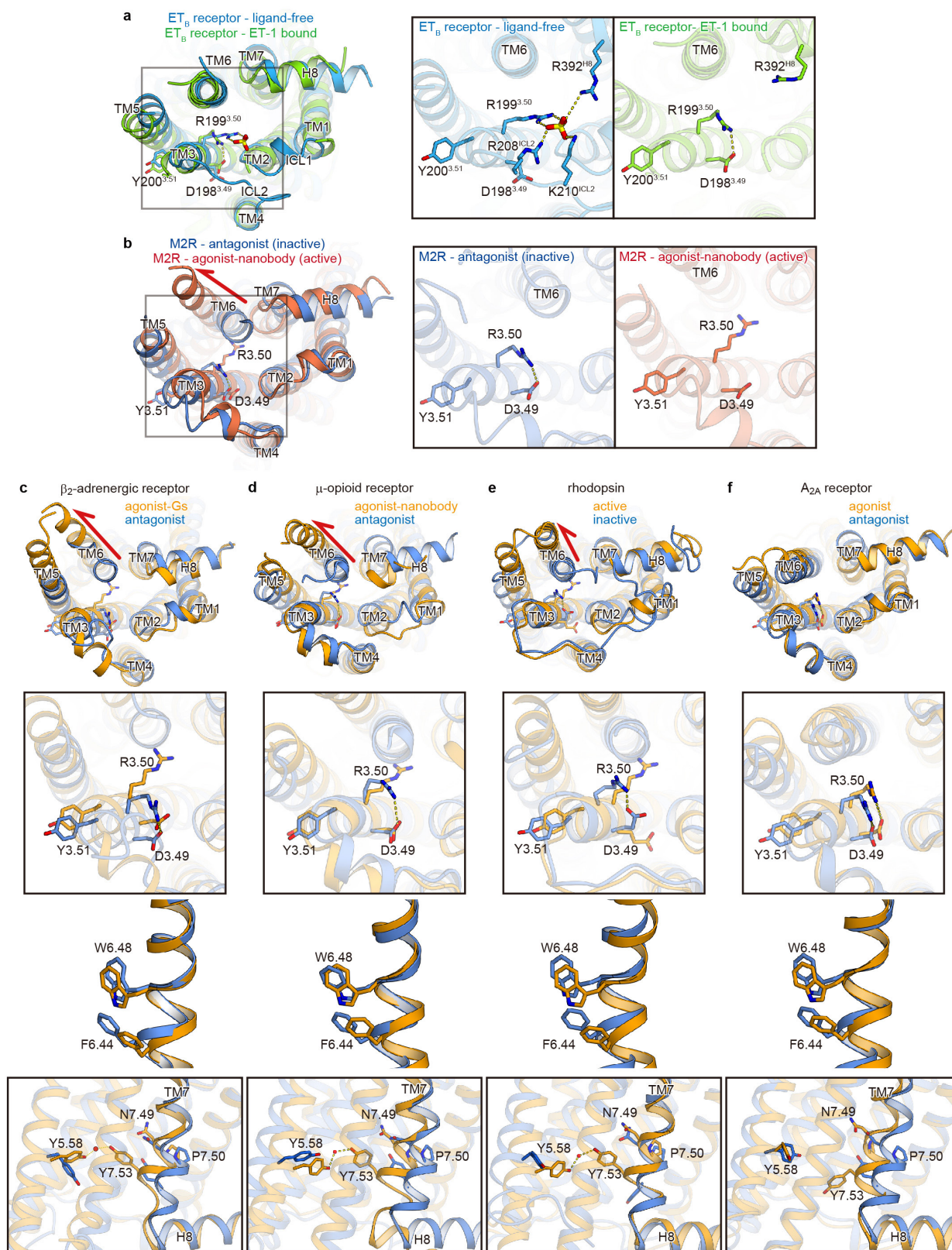
Extended Data Figure 7 | Competitive binding assays of mutant ET_B receptors. **a**, Competitive binding of ET-1 to wild-type and mutant ET_B receptors. The IC₅₀ values, representing the apparent half-maximal inhibitory concentration of ET-1 or ET-3 on ¹²⁵I-labelled ET-1 binding to mutant ET_B receptors, are indicated. The corresponding residues in ET_A are also indicated in the table, with the non-conserved residues represented in cyan. The letters a and b in the table indicate the host cells for expression: a, expressed in SF⁺ cell membranes; b, expressed in HEK293 cell membranes. **b**, **c**, Mutations that significantly affect ET-1 (**b**) and ET-3 (**c**) binding are indicated in the ET-1-bound structure. Mutated residues of ET_B are coloured according to the degree of decreased affinity, as shown in Extended Data Fig. 7a. ETs are indicated in ribbon representation with the colours as in **d**. **d**, Amino-acid sequences and

selectivities of the ET isopeptides. The residues different from ET-1 are highlighted in red. Intra-peptide disulfide bonds are indicated by yellow lines. **e**, Sequence conservation between human ET_A and ET_B receptors, mapped on the ET-1-bound ET_B structure. The residues of the receptor core are highly conserved between ET_A and ET_B receptors, and the amino-acid sequences of the C-terminal regions of the three endothelin isopeptides are identical, as shown in Fig. 2f, suggesting that the interactions between the C-terminal region of ET and the receptor core are conserved in any combination of ET isopeptides and receptor subtypes. However, the sequence conservation between ET_A and ET_B receptors suggests slightly divergent interactions through the extracellular portions of the receptors, including TM5, ECL1 and ECL2, where the N-terminal and α-helical regions of the ET-1 interact with the receptor.



Extended Data Figure 8 | Sequence conservation between ET_A and ET_B receptors. Amino-acid sequences of the thermostabilized crystallized construct (hET_BR-Y5), human ET_B (UniProt ID: P24530), rat ET_B (P26684), human ET_A (P25101) and rat ET_A (P26684) are aligned⁶⁵. Secondary structure elements for α-helices and β-strands are indicated by cylinders and arrows, respectively. Conservation of the residues between ET_A and ET_B is indicated as follows: red panels for completely conserved; red letters for partly conserved; and black letters for not conserved.

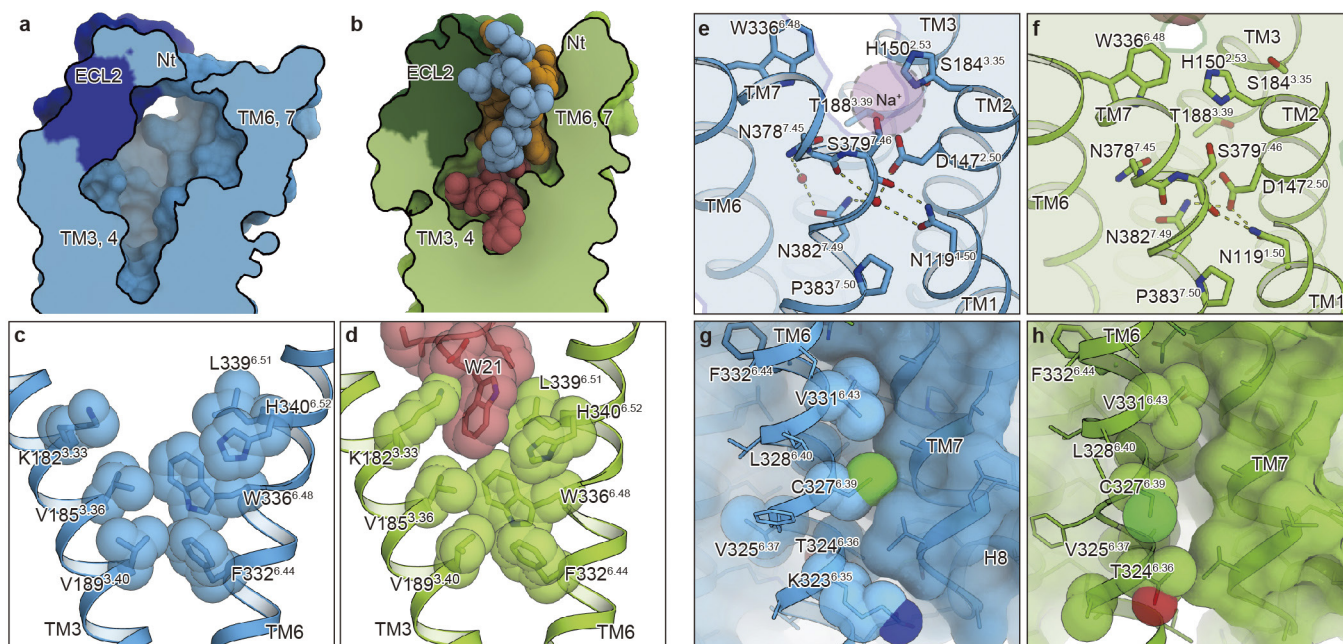
The thermostabilizing and the Cys-to-Ala mutations in the crystallized constructs are indicated with red and cyan letters, respectively. The residues with the Ballesteros-Weinstein number of X.50 in each TM helix are highlighted with yellow panels. The residues involved in the ET-1 binding are indicated by triangles, coloured according to the interacting regions of ET-1 (cyan, N-terminal region; orange, α-helical region; pink, C-terminal region).



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | Structural comparison of ET_B receptor and class A GPCRs. **a, b,** Cytoplasmic views of ligand-free and ET-1 bound ET_B receptors (**a**) and active and inactive M2R (**b**). The cytoplasmic architecture is similar between ligand-free and ET-1-bound ET_B receptors, while M2R shows the outward displacement of TM6 upon activation. Panels show close-up views of the E/DRY motif, with the important residues represented by sticks. The intra-helical salt bridge interaction is disrupted upon activation in M2R, and R3.50 points towards the centre of the receptor in the active conformation. Although the similar salt bridge formation is prevented by the sulfate ion in the ligand-free ET_B receptor, the rotamer orientations of R3.50 in both the ligand-free and ET-1 bound ET_B receptors represent the features of the inactive conformation. **c–f,** The structural comparisons of the β_2 -adrenergic receptor bound to an antagonist (PDB accession number 2RH1) and bound to an agonist and Gs (PDB accession number 3SN6) (**c**), the μ -opioid receptor bound to an antagonist (PDB accession number 4DKL) and bound to an agonist and nanobody (PDB accession number 5C1M) (**d**), rhodopsin in the ground

state (PDB accession number 3PXO) and in the active state (PDB accession number 2X72) (**e**), and the A_{2A} receptor bound to an antagonist (PDB accession number 4EIY) and bound to an agonist (PDB accession number 3QAK) (**f**). Cytoplasmic view (top), E/DRY motif on TM3 (middle upper), CWXP motif on TM6 (middle lower) and NPXXY motif on TM7 (bottom) of each receptor are shown. Red arrows in the upper panels indicate the outward displacement of TM6 that occurs upon receptor activation. The putative water molecule at the NPXXY motif in the β_2 -adrenergic receptor is represented by red circle. Residues involved in the structural rearrangement during receptor activation are represented by sticks, and hydrogen bonding interactions are indicated with yellow dotted lines. The agonist-bound A_{2A} receptor retains the structural features of the inactive conformation. The intra-helical salt bridge is formed in the E/DRY motif, and Tyr7.53 in the NPXXY motif is too far away to form a water-mediated hydrogen bonding interaction with Tyr5.58, although TM7 is shifted inwards (middle upper and lower panels in **f**).



Extended Data Figure 10 | ET-1-induced conformational changes in ET_B receptor. **a–d**, Cutaway representation and hydrophobic packing interaction in the receptor core for the ligand-free (**a**, **c**) and ET-1-bound (**b**, **d**) ET_B receptor. ET-1 binding induces the tightly packed hydrophobic core in the receptor. **e**, **f**, Collapse of the putative Na⁺ binding pocket in the ET_B receptor. Asp147^{2.50} and its surrounding residues are shown for the ligand-free (**e**) and ET-1-bound (**f**) ET_B receptor. Cross-sectional

representations of the orthosteric pocket are overlaid. The putative Na⁺ binding site is indicated with a purple-shaded circle. The electron density for the Na⁺ ion was not observed, probably because of the low resolution of the structure. **g**, **h**, TM6–7 interactions in the ligand-free (**g**) and ET-1-bound (**f**) ET_B receptors. TM1, TM2, TM3 and TM7 are shown as surface representations. The residues of TM6 directed towards TM7 are represented by CPK models.

Extended Data Table 1 | Crystallographic statistics of the ET-1-bound and ligand-free structures of the ET_B receptor

	ET-1 bound	ligand-free
Data collection	BL32XU	BL32XU
Space group	C222 ₁	C222 ₁
Number of crystals	1	7
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	73.0, 173.0, 109.4	74.0, 147.5, 107.8
α , β , γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Wavelength	1.0000	1.0000
Resolution (Å)	50-2.80 (2.96-2.80)	50-2.50 (2.59-2.50)
<i>R</i> _{meas}	0.160 (1.553)	0.252 (2.063)
<i>CC</i> _{1/2}	0.999 (0.877)	0.955 (0.535)
<i>I</i> / σ <i>I</i>	15.2 (1.6)	8.3 (1.1)
Completeness (%)	99.9 (100.0)	99.9 (99.8)
Redundancy	7.4 (7.5)	8.9 (8.0)
Refinement		
Resolution (Å)	50-2.80 (2.90-2.80)	50-2.50 (2.59-2.50)
No. reflections	17402 (1693)	20762 (2029)
<i>R</i> _{work} / <i>R</i> _{free}	0.238/0.278 (0.321/0.353)	0.250/0.281 (0.351/0.358)
No. atoms		
Protein	3639	3345
Water/Ion/Lipid	4	106
B-factors		
Protein	84.4	50.3
Water/Ion/Lipid	76.2	55.9
R.m.s. deviations		
Bond lengths (Å)	0.002	0.002
Bond angles (°)	0.49	0.45
Ramachandran plot		
Favored (%)	96.1	97.9
Allowed (%)	3.9	1.6
Disallowed (%)	0	0.5

*Values in parentheses are for highest-resolution shell.

m⁶A RNA methylation promotes *XIST*-mediated transcriptional repression

Deepak P. Patil¹, Chun-Kan Chen², Brian F. Pickering¹, Amy Chow², Constanza Jackson², Mitchell Guttman² & Samie R. Jaffrey¹

The long non-coding RNA X-inactive specific transcript (*XIST*) mediates the transcriptional silencing of genes on the X chromosome. Here we show that, in human cells, *XIST* is highly methylated with at least 78 N⁶-methyladenosine (m⁶A) residues—a reversible base modification of unknown function in long non-coding RNAs. We show that m⁶A formation in *XIST*, as well as in cellular mRNAs, is mediated by RNA-binding motif protein 15 (RBM15) and its paralogue RBM15B, which bind the m⁶A-methylation complex and recruit it to specific sites in RNA. This results in the methylation of adenosine nucleotides in adjacent m⁶A consensus motifs. Furthermore, we show that knockdown of *RBM15* and *RBM15B*, or knockdown of methyltransferase like 3 (*METTL3*), an m⁶A methyltransferase, impairs *XIST*-mediated gene silencing. A systematic comparison of m⁶A-binding proteins shows that YTH domain containing 1 (YTHDC1) preferentially recognizes m⁶A residues on *XIST* and is required for *XIST* function. Additionally, artificial tethering of YTHDC1 to *XIST* rescues *XIST*-mediated silencing upon loss of m⁶A. These data reveal a pathway of m⁶A formation and recognition required for *XIST*-mediated transcriptional repression.

XIST is a long non-coding RNA (lncRNA) that mediates the silencing of gene transcription on the X chromosome during female mammalian development¹ via the recruitment of specific protein complexes. These complexes have been identified in studies of the genetic domains involved in *XIST* silencing² as well as by recent unbiased proteomic screens that identified direct *XIST*-binding proteins using zero-distance ultraviolet irradiation-based crosslinking methods^{3,4}. Proteins that are bound directly or indirectly to *XIST* via protein intermediates have also been identified using crosslinking reagents such as formaldehyde⁵. These include HNRNPU (also known as SAF-A), which anchors *XIST* to the X chromosome⁶, SHARP (SPEN), which recruits HDAC3 (ref. 3), as well as PRC2, which introduces repressive chromatin marks⁷.

Here we show that *XIST*-mediated gene silencing requires adenosine methylation, a reversible RNA-modification pathway that forms m⁶A. Although the m⁶A modification is well-studied in mRNAs, m⁶A mapping studies⁸ have shown that m⁶A is also present in lncRNAs. Our data show that *XIST* is highly methylated and that m⁶A modifications are required for *XIST*-mediated gene silencing. Formation of m⁶A in *XIST* and mRNAs is mediated by two previously unknown components of the m⁶A methylation complex, RBM15 and RBM15B. These proteins bind and recruit the m⁶A-methylation complex to specific sites within *XIST*, leading to m⁶A formation at adjacent sites. Furthermore, we show that m⁶A in *XIST* recruits the m⁶A reader, YTHDC1 (hereafter DC1), and that the binding of DC1 to *XIST* promotes *XIST*-mediated gene repression. These studies reveal a role for m⁶A and DC1 as mediators of transcriptional repression via the lncRNA *XIST*.

RBM15 and RBM15B are required for gene silencing

Recent studies have shown that RBM15 binds to *XIST*^{3,4}. Previously, we found that the knockdown of *RBM15* did not block *XIST*-mediated gene silencing³; however, another study found the opposite to be true⁹. We therefore considered the possibility that another protein compensated for the function of RBM15 in our *RBM15* knockdown experiments. RBM15 possesses notable similarity to another protein, RBM15B, in sequence and domain organization, making it a suitable candidate for compensation of RBM15 function (Extended Data Fig. 1a).

To test the functional redundancy of these proteins, we first investigated whether RBM15 and RBM15B show similar binding patterns in *XIST* by mapping their binding sites using individual-nucleotide resolution UV crosslinking and immunoprecipitation (iCLIP)¹⁰ in human embryonic kidney 293T (HEK293T) cells. For all iCLIP experiments, we examined only the endogenous protein and identified antibodies that selectively precipitated each protein. We also confirmed that there was consistency between the transcriptome-wide iCLIP data set replicates (Extended Data Fig. 1b–g and Supplementary Tables 1, 2).

RBM15 and RBM15B showed a similar distribution of iCLIP tags (that is, processed reads; see Methods for further details) along the length of *XIST* (Fig. 1a and Extended Data Fig. 1h), including at the A-repeat region, an evolutionarily conserved region in the 5' region that is essential for the initiation of silencing¹¹. Additionally, RBM15 and RBM15B showed similar distributions of iCLIP tag clusters, which represent regions of enriched binding, and crosslinking-induced truncation sites (CITS), which represent direct contacts with *XIST* (Supplementary Tables 3, 4).

To assess whether RBM15 and RBM15B are required for *XIST*-mediated gene silencing, we used male mouse embryonic stem (ES) cells that express *Xist* on the X chromosome in a doxycycline-dependent manner³. *XIST*-mediated gene silencing is induced by 16 h of doxycycline (Dox)-induced *XIST* expression and is measured by quantifying the expression of two X-linked genes, *Gpc4* and *Atrx*, using single-molecule RNA fluorescence *in situ* hybridization (FISH)³. In these assays, we knocked down mRNAs using short interfering RNAs (siRNAs) and confirmed that each examined cell showed successful depletion of both the siRNA-targeted mRNA as well as Dox-induced *XIST* expression.

In wild-type siRNA-transfected cells, we observed the expected silencing of the X-linked genes. *Gpc4* transcript levels decreased from 21 copies (–Dox) to 1 copy (+Dox) per cell and *Atrx* transcript levels decreased from 17 to 1 copy per cell (Fig. 1b, c and Extended Data Fig. 2a, b). Knockdown of both *Rbm15* and *Rbm15b*, but not knockdown of either gene individually, prevented *XIST*-mediated gene silencing in these cells (Fig. 1b, c). This was also seen in a female mouse ES cell

¹Department of Pharmacology, Weill-Cornell Medical College, Cornell University, New York, New York 10065, USA. ²Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, California 91125, USA.

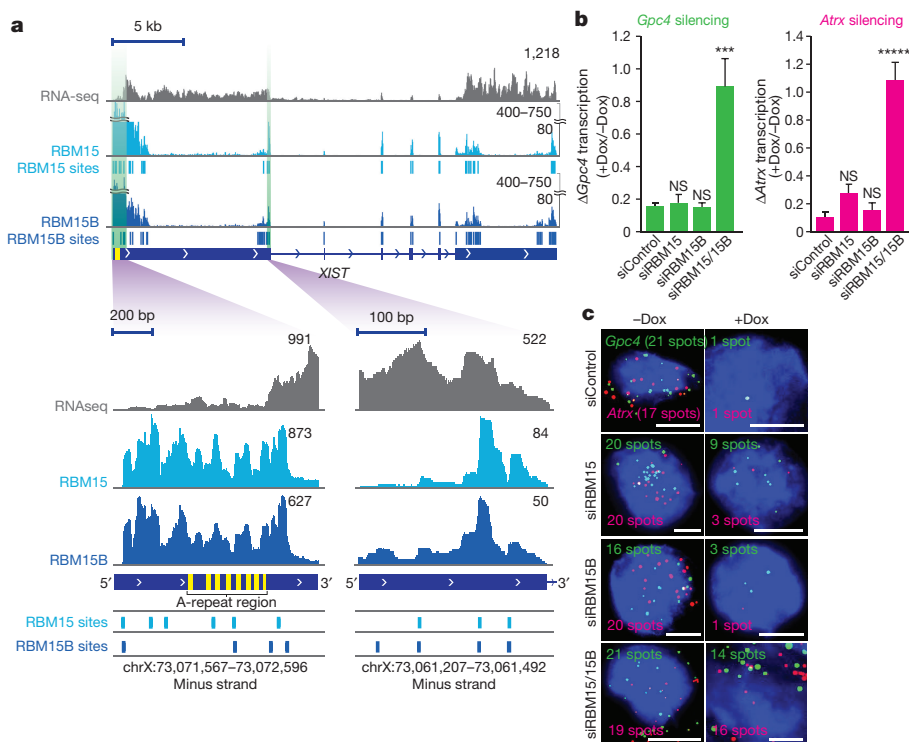


Figure 1 | RBM15 and RBM15B are necessary for *XIST*-mediated gene silencing. **a**, RBM15 and RBM15B show similar binding patterns in *XIST*. Shown is the distribution of normalized RBM15 and RBM15B iCLIP tags (in unique tags per million, uTPM) and statistically significant CITS. Light blue vertical lines, RBM15; dark blue vertical lines, RBM15B; $P < 0.0001$.

b, c, Knockdown of both *Rbm15* and *Rbm15b* (siRBM15/15B) impair *XIST*-mediated gene silencing. *XIST* expression was induced by doxycycline, and the X-linked genes *Gpc4* (green) and *Atrx* (red) were quantified by RNA-FISH (**b**). Representative FISH images are shown with DAPI nuclear counterstain (blue) (**c**). The number of detected RNA spots for both genes are indicated on each image. Scale bars, 5 μ m. Data are mean \pm s.e.m. for 50 cells from one experiment. *** $P < 0.001$, **** $P < 0.0001$, relative to siControl by unpaired two-sample *t*-test. NS, not significant.

line that similarly exhibits Dox-inducible *XIST* expression on one X chromosome (Extended Data Fig. 2c). RBM15 and RBM15B therefore have redundant function in mediating *XIST*-mediated transcriptional silencing.

RBM15/RBM15B link the methylation complex to *XIST*

RBM15 and RBM15B were recently identified as high-confidence interactors with Wilms tumour-associated protein (WTAP) in a proteomic analysis¹². WTAP binds METTL3 (refs 13–15), the methyltransferase that mediates methylation of m⁶A in mRNA¹⁶, and is recruited to RNAs via an unknown adaptor protein to trigger m⁶A formation¹⁴.

We therefore investigated whether RBM15 and/or RBM15B is a component of the WTAP–METTL3 complex, targeting it to RNA. Immunoprecipitation of RBM15 or RBM15B from HEK293T nuclear lysates co-precipitated METTL3 (Fig. 2a). Knockdown of WTAP reduced the interaction between METTL3 and both RBM15 and RBM15B (Fig. 2a), indicating that this interaction is mediated by WTAP. A reciprocal immunoprecipitation similarly indicated that METTL3 binds RBM15 and RBM15B in a WTAP-dependent manner (Extended Data Fig. 3a–c).

To determine whether both RBM15 and RBM15B (RBM15/15B) can recruit WTAP–METTL3 to *XIST*, we treated HEK293T cells with formaldehyde to crosslink *XIST* to any bound proteins. We then immunoprecipitated METTL3 from the cell lysates and measured the amount of bound *XIST* by quantitative reverse transcription PCR (qRT–PCR) at regions with and without RBM15/15B-binding sites. METTL3 immunoprecipitates contained significantly higher levels of *XIST* than control immunoprecipitates at these binding sites (Fig. 2b and Extended Data Fig. 3d, e). This interaction was impaired after knockdown of WTAP, RBM15, and/or RBM15B, with the greatest loss following knockdown of both RBM15 and RBM15B double knockdown (Fig. 2b). This led us to believe that RBM15/15B is the component of the methylation complex that accounts for its recruitment to *XIST*.

RNA-anchored methylation complexes

Our initial m⁶A mapping studies, using methylated RNA immunoprecipitation followed by sequencing (MeRIP–seq), showed *XIST* contained m⁶A modifications⁸, although this approach was at low

resolution. More recently, we mapped m⁶A at single-nucleotide resolution using m⁶A iCLIP (miCLIP)¹⁷. Analysis of the miCLIP data set shows 78 putative m⁶A residues in *XIST*, some of which are localized at or near the A-repeat region (Fig. 3a and Extended Data Fig. 4a). To investigate whether RBM15 and RBM15B mediate m⁶A formation in *XIST*, we measured m⁶A levels in *XIST* in wild-type control and RBM15/15B-deficient cells. Methylated *XIST* was precipitated with an m⁶A-specific antibody and *XIST* levels were quantified from three m⁶A-containing regions (Fig. 3a). Knockdown of *METTL3*, *RBM15*, *RBM15B*, and both *RBM15* and *RBM15B* resulted in significantly reduced levels of methylated *XIST*, with the largest reduction in m⁶A levels following *RBM15/RBM15B* double knockdown (Fig. 3b and Extended Data Fig. 3d–f). This indicates that RBM15 and RBM15B promote *XIST* methylation by recruiting WTAP–METTL3.

We observed that m⁶A residues are typically located in the vicinity of RBM15 and RBM15B iCLIP clusters on *XIST* (Extended Data Fig. 4b). Indeed, the median distance between each RBM15 or RBM15B CITS in *XIST* and the closest m⁶A was 45 or 28.5 nucleotides, respectively (Extended Data Fig. 4c). By contrast, the distance between m⁶A and randomly picked sites along *XIST* was approximately 70–90 nucleotides ($P = 0.0026$, RBM15; $P = 0.0001$, RBM15B). Thus, m⁶A residues are positioned significantly closer to RBM15 and RBM15B sites than would be expected by chance. This proximity suggests that RBM15/15B recruits the WTAP–METTL3 complex to methylate adenosine bases that lie in proximal m⁶A consensus sites.

We next asked whether RBM15/15B binds next to m⁶A bases in mRNA. Using our single-nucleotide-resolution m⁶A data set in mRNA¹⁷, we calculated the spatial relationship of RBM15/15B-binding sites relative to m⁶A residues. As a control, we measured the binding of RBM15 and RBM15B relative to non-methylated adenosines that fall within the m⁶A consensus DRACH sequence (where D denotes A/G/U, R denotes A/G and H denotes A/C/U)¹⁷. These sites lack miCLIP reads and thus are non-methylated. Transcriptome-wide analysis shows that RBM15/15B-binding sites are significantly enriched on either side of m⁶A residues, while minimal enrichment is seen at the nearest non-methylated DRACH site (Extended Data Fig. 5a). RBM15/15B-binding sites are characterized by U-rich motifs (Extended Data Fig. 5c–e) that are readily detected adjacent to m⁶A residues on individual transcripts (Extended Data Fig. 5b).

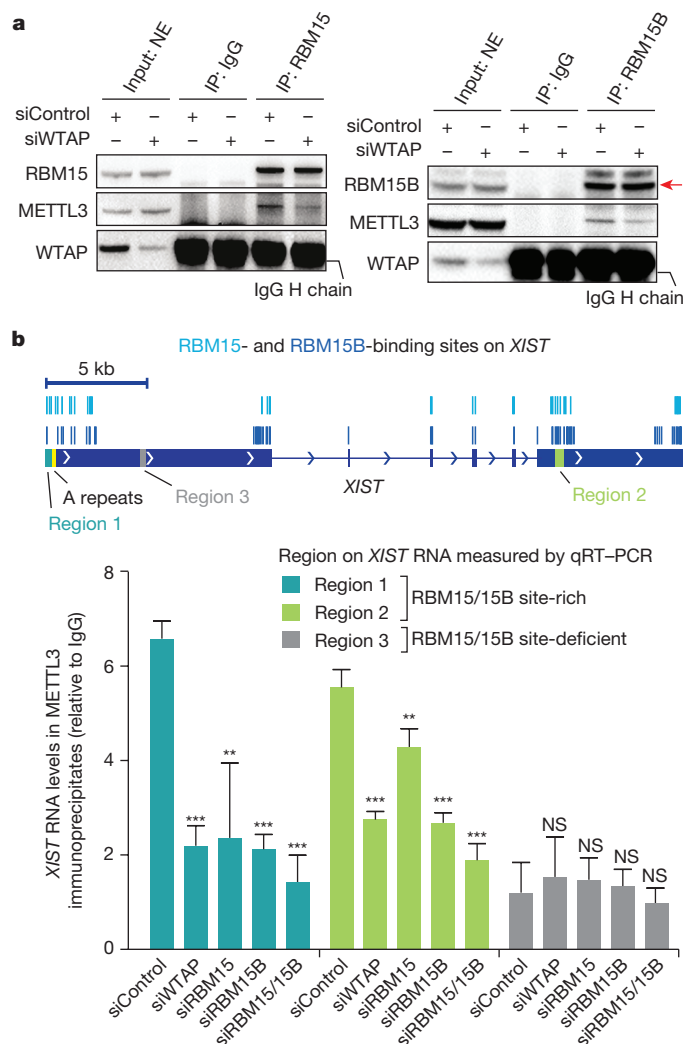


Figure 2 | RBM15 and RBM15B recruit METTL3 to *XIST*. **a**, RBM15 and RBM15B interact with METTL3 in a WTAP-dependent manner. RBM15 (left) and RBM15B (right) were immunoprecipitated from HEK293T nuclear extracts. Co-immunoprecipitation of METTL3 was reduced in siWTAP-transfected cells. The IgG heavy chain (H chain) prevents visualization of WTAP; however, knockdown is seen in the input sample. NE, nuclear extracts. **b**, Quantification of METTL3-bound *XIST* upon knockdown of methylation machinery components. *XIST* was quantified by qRT-PCR using regions selected based on the presence or absence of RBM15- and RBM15B-binding sites (indicated with light blue and blue lines, respectively). Data are mean \pm s.e.m. from three independent experiments. ** $P < 0.001$, *** $P < 0.0001$, relative to siControl by unpaired two-sample *t*-test.

Notably, knockdown of both *RBM15* and *RBM15B* resulted in a substantial drop in m^6A levels in poly(A) RNA (Extended Data Fig. 5f, g), indicating that RBM15 and RBM15B direct methylation of adenosine residues at sites in both mRNA and *XIST*.

XIST m^6A is required for gene silencing

XIST has more mapped m^6A residues than any other RNA (Supplementary Tables 5, 6), raising the possibility that m^6A may mediate important aspects of *XIST* function. The role of m^6A in *XIST*-mediated gene silencing cannot be tested in *Mettl3*^{-/-} mouse ES cells because these cells do not express *XIST* owing to the persistent expression of *XIST*-suppressing pluripotency genes¹⁸. We thus used the Dox-inducible *XIST*-expression system to assess the role of METTL3 in *XIST*-mediated transcriptional silencing. *METTL3* knockdown reduces m^6A levels across the transcriptome, including in *XIST*¹⁹. In control siRNA-transfected cells, we observed the expected silencing

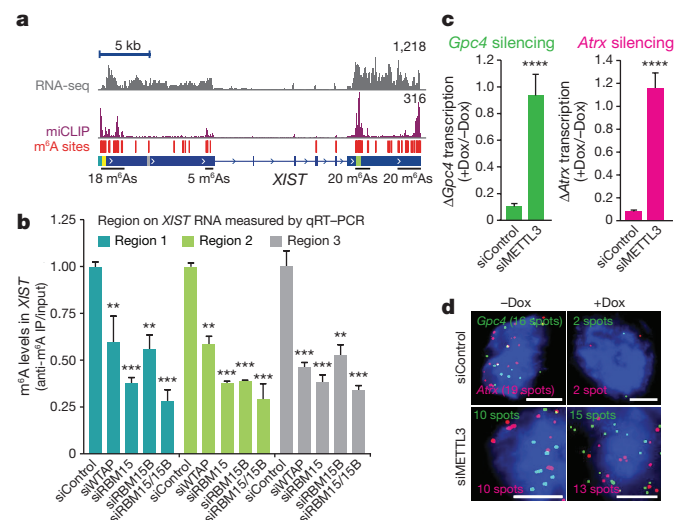


Figure 3 | N^6 -adenosine methylation is necessary for *XIST*-mediated gene silencing. **a**, m^6A residues (red lines) identified via miCLIP are broadly distributed along *XIST*. Normalized miCLIP¹⁷ tags are shown in purple. **b**, Methylation of *XIST* requires RBM15 and RBM15B. m^6A levels in *XIST* were quantified by m^6A -RNA immunoprecipitation followed by qRT-PCR of three m^6A regions of *XIST*. Data are mean \pm s.e.m. from six samples coming from three technical replicates of two biological replicates. *** $P < 0.0001$, ** $P < 0.001$ relative to siControl by unpaired two-sample *t*-test. **c**, **d**, m^6A promotes *XIST*-mediated gene silencing. *XIST* expression was induced by Dox, and X-linked genes *Gpc4* (green) and *Atrx* (red) were quantified by RNA-FISH (**c**). Representative FISH images are shown (**d**). The number of detected RNA spots is indicated on each image. Scale bars, 5 μ m. Data are mean \pm s.e.m. for 50 cells from one experiment. **** $P < 0.005$ relative to siControl by an unpaired two-sample *t*-test.

of X-linked genes upon *XIST* induction (Fig. 3c, d and Extended Data Fig. 2c–e). However, in siMetl3-treated cells, *XIST* was induced but failed to silence *Gpc4* and *Atrx* expression (Fig. 3c, d and Extended Data Fig. 2d, e). A similar silencing defect was seen in a female mouse ES cell line with Dox-inducible *XIST* expression (Extended Data Fig. 2c). Therefore, m^6A is required for *XIST*-mediated transcriptional silencing.

DC1 binds *XIST* to mediate gene silencing

We next investigated the mechanism by which m^6A in *XIST* is recognized in order to mediate transcriptional silencing. m^6A residues are recognized by the YTH proteins²⁰ which comprise three members of the YTHDF family (DF1, DF2, and DF3), YTHDC1 (DC1) and YTHDC2 (DC2) (Extended Data Fig. 6a). DF1, DF2, DF3 and DC2 are primarily cytoplasmic^{21–24}, whereas DC1 is located primarily in the nucleus²⁴.

Using iCLIP, we assessed the transcriptome-wide binding properties of the endogenous YTH proteins and determined whether any interacted preferentially with m^6A in *XIST* (Extended Data Figs 6, 7 and Supplementary Table 1, 2). In this analysis, we quantified the binding of YTH proteins at each of the 78 mapped m^6A residues in *XIST* as well as the other 11,452 mapped m^6A residues in the transcriptome. Each m^6A residue was assigned an intensity value that was defined as the normalized number of miCLIP tags for each m^6A residue¹⁷. This value is influenced by both the transcript abundance and the m^6A stoichiometry. Next, the binding of each YTH protein to each m^6A residue was determined using the normalized number of mapped iCLIP tags at the m^6A site. For most m^6A residues, the miCLIP intensity value increased with the amount of bound YTH protein (Fig. 4a); however, only DC1 showed clear preferential binding for *XIST* m^6A residues (Fig. 4a, b and Extended Data Fig. 8a–c).

A direct comparison of iCLIP tags on *XIST* also showed that DC1 is the only YTH protein to exhibit prominent *XIST* binding (Fig. 4c and Extended Data Fig. 8a–d). Notably, the DC1 iCLIP tag clusters overlap

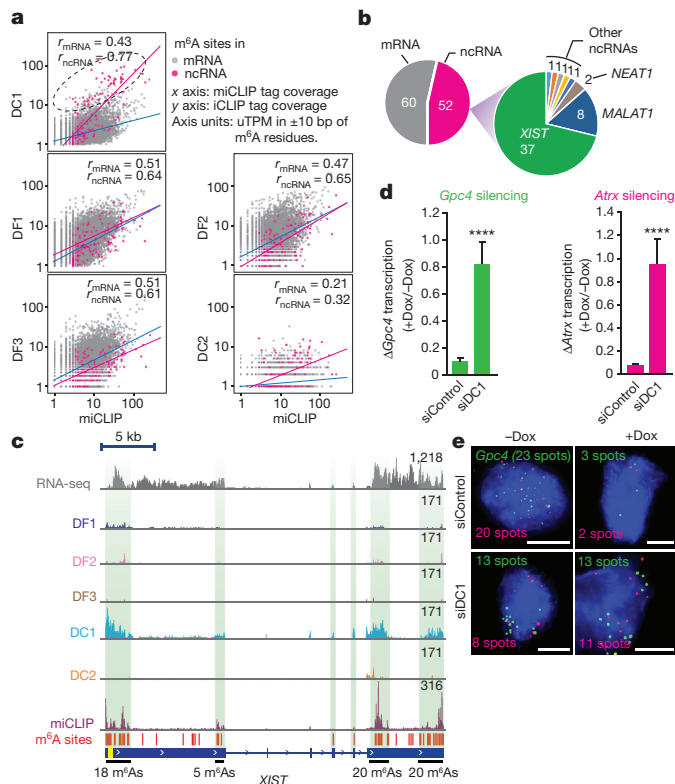


Figure 4 | DC1 binds *XIST* m⁶A residues and promotes *XIST*-mediated gene silencing. **a**, YTH iCLIP tag coverage at 11,530 annotated m⁶A residues. Correlation coefficients for mRNA m⁶A (grey) and non-coding RNA (ncRNA) m⁶A (magenta) are indicated. DF1, DF2 and DF3 show similar correlations between m⁶A abundance and YTH binding for mRNAs (blue line) and ncRNAs (magenta line). DC1 shows preference for ncRNA m⁶A, with the top 1% of DC1-bound m⁶A indicated (dotted ellipse). **b**, mRNA/ncRNA distribution of the top 1% of DC1-bound m⁶A sites. Most detected ncRNA m⁶A are present on *XIST* (indicated in green). **c**, Normalized tag distributions for each YTH protein on *XIST* shows predominantly DC1 binding. High-density m⁶A regions are indicated by green shading. **d, e**, *Ythdc1* knockdown (siDC1) impairs *XIST*-mediated gene silencing. *XIST* was induced by Dox, and X-linked genes *Gpc4* (green) and *Atrx* (red) were quantified by RNA-FISH. Representative FISH images are shown (**e**). The number of detected RNA spots is indicated on each image. Scale bars, 5 μm. Data are mean ± s.e.m. *****P* < 0.005 relative to siControl by an unpaired two-sample *t*-test.

with the *XIST* m⁶A miCLIP tag clusters, consistent with the binding of DC1 to m⁶A residues in *XIST* (Fig. 4c and Extended Data Fig. 8d).

The binding of DC1 to *XIST* could also be confirmed through the co-immunoprecipitation of DC1 and *XIST* using antibodies against DC1, with *XIST* detected by qRT-PCR using primers that detect either of the two regions with a high DC1 iCLIP signal (Extended Data Fig. 9a). *XIST* pulldown was reduced following the knockdown of methylation machinery components (*METTL3*, *WTAP*, *RBM15*, *RBM15B*, and *RBM15* and *RBM15B* double knockdown). Furthermore, DC1 was enriched in the *XIST* nuclear subcompartment in comparison to autosomal domains as measured by 3D structured illumination super-resolution microscopy (3D-SIM) (Extended Data Fig. 9b–d). This localization was reduced following knockdown of *METTL3* or both *RBM15* and *RBM15B* (Extended Data Fig. 9e). Together, these data show that DC1 binds to *XIST* in an m⁶A-dependent manner.

We then assessed whether DC1 is required for *XIST*-mediated transcriptional silencing. Knockdown of *DC1* but not *DF1*, *DF2*, *DF3* or *DC2* prevented *XIST*-mediated gene silencing in cells with Dox-induced *XIST* expression (Fig. 4d, e and Extended Data Fig. 2f–j) and in differentiating female mouse ES cells (Extended Data Fig. 2i). To determine whether DC1 binding to *XIST* promotes *XIST*-mediated gene silencing,

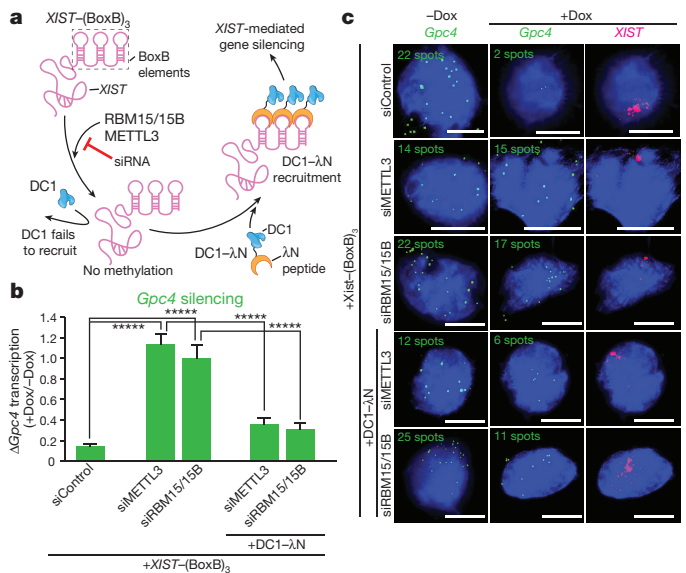


Figure 5 | m⁶A-independent tethering of DC1 to *XIST* is sufficient to exert *XIST*-mediated gene silencing. **a**, Schematic of tethering approach. The 3' end of *XIST* was genomically modified with three BoxB sequences (*XIST*-(BoxB)₃). m⁶A-dependent recruitment of DC1 is blocked in methylation-deficient cells; however, artificial tethering can be achieved with DC1-λN, which binds to the BoxB elements in *XIST*-(BoxB)₃. **b, c**, Dox-induced expression of *XIST*-(BoxB)₃ results in gene silencing in siControl-transfected cells, but not in siMETTL3 or siRBM15 and siRBM15B co-transfected cells. DC1-λN rescued silencing in these cells, suggesting that the primary function of m⁶A in *XIST*-mediated gene silencing is to recruit DC1 to *XIST*. Quantification of *Gpc4* expression is shown in **b**. Representative FISH images showing DAPI-stained nuclei (blue), *Gpc4* RNA (green), and *XIST* (pink) are shown in **c**. Scale bars, 5 μm. Data are mean ± s.e.m. in **b** for 50 cells from one experiment. *****P* < 0.0001 by unpaired two-sample *t*-test.

we tethered DC1 to *XIST* using an *XIST* transcript with three BoxB hairpins appended to the 3' end (*XIST*-(BoxB)₃) (Fig. 5a). These hairpins bind the λN peptide fused to the C terminus of DC1, allowing the BoxB hairpins to bind the λN peptide. Dox-induced expression of *XIST*-(BoxB)₃ caused transcriptional repression of *Gpc4* and this silencing was lost following knockdown of *Mettl3* or both *Rbm15* and *Rbm15b* (Fig. 5b, c). However, *XIST*-mediated gene silencing was rescued when DC1-λN was expressed (Fig. 5b, c). Thus, recruitment of DC1 to *XIST* is sufficient to induce its repressive function in the absence of the methylation machinery. Taken together, these data suggest that m⁶A methylation of *XIST* triggers binding to DC1, which promotes *XIST*-mediated transcriptional silencing.

Discussion

Although the m⁶A modification has been well-characterized in mRNA, no function for m⁶A in lncRNAs has previously been demonstrated. Here we show that m⁶A functions to enable the transcriptional repression effects of *XIST*. *XIST* is highly enriched in m⁶A throughout its length, enabling the recruitment of the nuclear m⁶A binding protein DC1. The importance of m⁶A in *XIST* function is highlighted by the fact that diverse components of the m⁶A methylation complex bind *XIST* and are required for *XIST*-mediated gene silencing. Together, these discoveries reveal a role for RNA modification in lncRNA function and describe the assembly of *XIST* into a transcriptionally repressive ribonucleoprotein complex (Extended Data Fig. 10a).

Recent proteomic studies have revealed large numbers of *XIST*-binding proteins^{3–5,9}, several of which we now recognize as contributing to m⁶A formation or recognition. For example, WTAP was identified in a proteomic analysis of *XIST*-associated proteins⁵ and was shown to be required for *XIST*-mediated gene silencing in a functional screen⁹. Although WTAP has numerous functions, our data support the idea

that its m⁶A methylation-promoting effects are required for *XIST*-mediated gene silencing. DC1 was also observed in a proteomic analysis of formaldehyde-crosslinked proteins bound to *XIST*⁵.

Similarly, RBM15 was shown to be required for *XIST*-mediated gene silencing⁹ and was also identified as an *XIST*-binding protein^{3,5,9}. Our data suggest that RBM15/15B is a component of the m⁶A methylation complex that binds *XIST*, and that it is this methylation role that is essential in bringing about the silencing defect observed when both are knocked down. RBM15 and RBM15B appear to have redundant functions as both need to be knocked down in order to deplete m⁶A to sufficient levels to impair *XIST* function. The large number of m⁶A residues in *XIST* ensures that at least a few will bind to DC1 to activate gene-silencing mechanisms.

The identification of the WTAP–METTL3 complex¹³ and its role in m⁶A formation¹⁴ raised several important questions. First, why are some RNAs methylated, while others lack m⁶A? Second, why are only a subset of DRACH-site adenosine residues selected for methylation, despite the high prevalence of DRACH consensus sites in RNA²⁵? Our data sheds light on these questions. RBM15 and RBM15B, proteins that associate with WTAP–METTL3 and contain RNA-binding domains, enable the binding of WTAP–METTL3 to specific mRNAs, as well as *XIST*. The localized binding at specific sites in the RNA sequence allows for the selective methylation of adjacent DRACH sites while leaving distant DRACH sites unmethylated. The three-dimensional RNA structure of *XIST* could promote further adenosine methylation by bringing distant DRACH consensus sites into the proximity of the RBM15/15B-anchored methylation complex.

Our single-nucleotide-resolution map of m⁶A (ref. 17) showed that RBM15/15B is found adjacent to methylated but not non-methylated DRACH sequences in the mRNA transcriptome. The double knock-down of *RBM15* and *RBM15B* markedly reduce m⁶A levels in mRNA, supporting the idea that RBM15/15B-binding determines which DRACH sites are methylated in the transcriptome.

How DC1 binding to *XIST* leads to gene silencing remains unclear. However, a recent proteomics study exploring DC1 binding partners²⁷ may provide initial mechanistic insights. These partners include SHARP, LBR, HNRNPU, and HNRNPK which each have distinct roles in the initiation of transcriptional silencing (Extended Data Fig. 10b–e). Analysis of the DC1 interaction network, based on an independent protein–protein interaction database²⁸, also identifies additional interactions with components of the PRC1 and PRC2 complexes (Extended Data Fig. 10b–e and Supplementary Table 7). Various *XIST*-interacting gene-silencing proteins may bind to DC1 and utilize the ability of DC1 to bind m⁶A residues on *XIST* to achieve additional specificity in the binding of precise locations on *XIST*. Further experiments are required both to determine whether DC1 directly affects binding of these silencing proteins and to explore the mechanisms used by DC1 to enable m⁶A-dependent transcriptional silencing.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 February; accepted 29 July 2016.

Published online 7 September 2016.

- Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. Requirement for *Xist* in X-chromosome inactivation. *Nature* **379**, 131–137 (1996).
- Wutz, A. Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. *Nat. Rev. Genet.* **12**, 542–553 (2011).
- McHugh, C. A. *et al.* The *Xist* lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232–236 (2015).
- Minajigi, A. *et al.* Chromosomes. A comprehensive *Xist* interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* **349**, aab2276 (2015).
- Chu, C. *et al.* Systematic discovery of *Xist* RNA binding proteins. *Cell* **161**, 404–416 (2015).
- Hasegawa, Y. *et al.* The matrix protein hnRNP U is required for chromosomal localization of *Xist* RNA. *Dev. Cell* **19**, 469–476 (2010). 10.1016/j.devcel.2010.08.006

- Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750–756 (2008).
- Meyer, K. D. *et al.* Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149**, 1635–1646 (2012).
- Moindrot, B. *et al.* A pooled shRNA screen identifies *Rbm15*, *Spen*, and *Wtap* as factors required for *Xist* RNA-mediated silencing. *Cell Reports* **12**, 562–572 (2015).
- Köznig, J. *et al.* iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* **17**, 909–915 (2010).
- Wutz, A., Rasmussen, T. P. & Jaenisch, R. Chromosomal silencing and localization are mediated by different domains of *Xist* RNA. *Nat. Genet.* **30**, 167–174 (2002).
- Horiuchi, K. *et al.* Identification of Wilms' tumor 1-associating protein complex and its role in alternative splicing and the cell cycle. *J. Biol. Chem.* **288**, 33292–33302 (2013).
- Zhong, S. *et al.* MTA is an *Arabidopsis* messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. *Plant Cell* **20**, 1278–1288 (2008).
- Agarwal, S. D., Blitzblau, H. G., Hochwagen, A. & Fink, G. R. RNA methylation by the MIS complex regulates a cell fate decision in yeast. *PLoS Genet.* **8**, e1002732 (2012).
- Ping, X. L. *et al.* Mammalian WTAP is a regulatory subunit of the RNA N⁶-methyladenosine methyltransferase. *Cell Res.* **24**, 177–189 (2014).
- Bokar, J. A., Rath-Shambaugh, M. E., Ludwiczak, R., Narayan, P. & Rottman, F. Characterization and partial purification of mRNA N⁶-adenosine methyltransferase from HeLa cell nuclei. Internal mRNA methylation requires a multisubunit complex. *J. Biol. Chem.* **269**, 17697–17704 (1994).
- Linder, B. *et al.* Single-nucleotide-resolution mapping of m⁶A and m⁶Am throughout the transcriptome. *Nat. Methods* **12**, 767–772 (2015).
- Plath, K. & Lowry, W. E. Progress in understanding reprogramming to the induced pluripotent state. *Nat. Rev. Genet.* **12**, 253–265 (2011).
- Schwartz, S. *et al.* Perturbation of m⁶A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. *Cell Reports* **8**, 284–296 (2014).
- Dominissini, D. *et al.* Topology of the human and mouse m⁶A RNA methylomes revealed by m⁶A-seq. *Nature* **485**, 201–206 (2012).
- Wang, X. *et al.* N⁶-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117–120 (2014).
- Wang, X. *et al.* N⁶-methyladenosine modulates messenger RNA translation efficiency. *Cell* **161**, 1388–1399 (2015).
- Morohashi, K. *et al.* Cyclosporin A associated helicase-like protein facilitates the association of hepatitis C virus RNA polymerase with its cellular cyclophilin B. *PLoS One* **6**, e18285 (2011).
- Berglund, L. *et al.* A gene-centric Human Protein Atlas for expression profiles based on antibodies. *Mol. Cell. Proteomics* **7**, 2019–2027 (2008).
- Meyer, K. D. & Jaffrey, S. R. The dynamic epitranscriptome: N⁶-methyladenosine and gene expression control. *Nat. Rev. Mol. Cell Biol.* **15**, 313–326 (2014).
- Kwon, S. C. *et al.* The RNA-binding protein repertoire of embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1122–1130 (2013). 10.1038/nsmb.2638
- Xiao, W. *et al.* Nuclear m⁶A reader YTHDC1 regulates mRNA splicing. *Mol. Cell* **61**, 507–519 (2016). 10.1016/j.molcel.2016.01.012
- Cowley, M. J. *et al.* PINA v2.0: mining interactome modules. *Nucleic Acids Res.* **40**, D862–D865 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank members of the Jaffrey and Guttman laboratories for comments and suggestions, A. Olarerin-George for discussions on computational analysis, J. -Y. Kim for discussions on immunostaining and imaging, the members of the Weill Cornell Epigenomics Core for their assistance in high-throughput sequencing, A. North and other staff members of the Bio-Imaging Resource Center at the Rockefeller University for their assistance with 3D-SIM imaging, and S. Mukherjee for input on image analysis. This work was supported by NIH grants R01CA186702 (to S.R.J.); T32CA062948 (B.F.P.); T32GM07616 (C.-K.C.) and an NIH Director's Early Independence Award (DP5OD012190), the Rose Hills Foundation, Edward Mallinckrodt Foundation, Sontag Foundation, Searle Scholars Program, the Pew-Stewart Scholars program, and funds from the California Institute of Technology (M.G.).

Author Contributions All authors designed the experiments and analysed data. D.P.P., B.F.P. and S.R.J. primarily oversaw iCLIP, biochemistry, and SIM experiments; C.-K.C. and M.G. oversaw experiments on *XIST*-mediated silencing in mouse ES cells. D.P.P. performed the iCLIP experiments, collected data, performed 3D-SIM imaging, analysed data, and prepared figures. C.-K.C. and C.J. performed *XIST* mouse ES experiments, collected data, and prepared figures. A.C. made and characterized the female pSM33 ES cell line. D.P.P. and S.R.J. wrote the manuscript and all authors provided input and comments.

Author Information iCLIP datasets generated in this study have been deposited in the Gene Expression Omnibus database under accession number GSE78030. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.R.J. (srj2003@med.cornell.edu).

Reviewer Information *Nature* thanks S. Schwartz and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Cell culture. HEK293T/17 (ATCC CRL-11268) cells were maintained in $1 \times$ DMEM (11995-065, Life Technologies) with 10% FBS, 100 U ml^{-1} penicillin and $100 \mu\text{g ml}^{-1}$ of streptomycin under standard tissue culture conditions. Cells were split using TrypLE Express (Life Technologies) according to manufacturer's instructions. Mouse ES cells expressing *Xist* RNA from the endogenous locus under a Tet-driven promoter (pSM33 ES cell line) were maintained as previously described²⁹. Cell lines were not tested for mycoplasma contamination.

Generation of female pSM33 cell line. The Tet-regulated promoter was inserted at the promoter region of the endogenous *Xist* locus of mouse female ES cell line (F₁ 2-1 line, derived from a $129 \times$ castaneous F₁ mouse cross) using CRISPR-mediated homologous recombination. Clonal cell lines derived from single cells were screened for the presence of Tet-inducible promoter by PCR. Promoter integration was confirmed by Sanger sequencing with primers flanking the insertion site. Recombinant *Xist* alleles were further identified by SNP analysis. A clonal line with promoter insertion in the 129 allele was used for studying *Xist*-mediated gene silencing.

Insertion of BoxB sequence elements in *Xist*. Three BoxB sequence elements were inserted at the 3' end of the endogenous *Xist* loci in the male pSM33 cell line using CRISPR-mediated homologous recombination. In brief, cells were first co-transfected with a plasmid expressing Cas9 under a CAG promoter, a short guide RNA (Target sequence: 5'-CCTCATCTCATGTCTTCTC-3'), and a ssDNA ultramer (IDT) containing three BoxB elements (5'-GGGCCCTGAAGAAGGGCCCATGGGCCCTGAAGAAGGGCCCATAGGGCCCTGAAGAAGGGCCCATAGGGCCCTGAAGAAGGGCC-3'; underlined bases mark the BoxB sequence) flanked by 70-nucleotide-long DNA sequence identical to the upstream and downstream genomic DNA sequence at the point of BoxB insertion. Cells were sorted and single colonies were screened for the insertion of BoxB elements by PCR. Insertion was further confirmed by Sanger sequencing. Recombinant clones were tested for X-chromosome silencing by induction of *Xist* expression and *Gpc4* and *Atrx* RNA-FISH. A clone showing silencing identical to the non-recombinant cell line was used for DC1- λ N-*XIST* tethering functional assay.

Construction of λ N-3 \times Flag epitope-tagged DC1 expression construct. A human YTHDC1-encoding open reading frame (ORF) was PCR-amplified from oligo-(dT)₁₈-primed HEK293T cDNA using hYTHDC1-EcoRI-F and hYTHDC1-XhoI-R primers (Supplementary Table 8). The PCR fragment was initially cloned in pcDNA3-Flag-HA (1436 pcDNA3-Flag-HA was a gift from W. Sellers; Addgene plasmid 10792) plasmid at EcoRI and XhoI sites. Full-length YTHDC1 was then PCR amplified and subcloned into pCAG-GW- λ N-3 \times Flag-BSD construct using the Gateway entry cloning system (Invitrogen). This plasmid (pCAG-GW-hYTHDC1- λ N-3 \times Flag-BSD) expresses human YTHDC1 protein with a C-terminal λ N-3 \times Flag tag under CAG promoter. We verified that λ N-3 \times Flag-tagged DC1 protein was still functional by ensuring that it could rescue knockdown of the endogenous protein.

Generation of *Ythdc1*^{+/-} female ES cells. *Ythdc1*^{+/-} female ES cell line was generated using the CRISPR-Cas9 system. In brief, female ES cells were co-transfected with a Cas9-expressing pCAG plasmid and a pool of short guide RNAs targeting the region around the first codon of the *Ythdc1* ORF at the endogenous loci to generate frameshift mutations causing disruption in the reading frame. Target DNA sequences were 5'-AAGCCGAGGGCAGCCATGG-3', 5'-GCGGTGGCGGCGCGGAAGC-3' and 5'-CGGCGGAAGCCGAGG GCAG-3'. We screened 24 colonies derived from single cells for the presence of frame-shift mutations at the desired location in *Ythdc1* gene using PCR and Sanger sequencing, with primers flanking the target site. No clone showed a homozygous frame-shift mutation, suggesting that homozygous *Ythdc1* deletion is lethal. Only clones with heterozygous frame shift mutations were detected. Confirmation of the presence of a heterozygous knockout of *Ythdc1* (*Ythdc1*^{+/-}) was performed by RNA-FISH and immunofluorescence. A clonal cell line showing a 50% reduction in the expression level of *Ythdc1* mRNA and protein were used for assaying X-chromosome silencing.

Antibodies. Details of the antibodies used in this study are given in Supplementary Table 1.

siRNA and shRNA transfection. Target sequences of siRNA and short hairpin RNA (shRNA) used in this study are listed in Supplementary Table 9. For validation of antibodies for iCLIP, 20 nM siRNA was transfected using Pepmute transfection reagent (Signagen) and pSuperior-EGFP shRNA plasmid (OligoEngine) was transfected using Fugene HD transfection reagent (Promega) according to the manufacturer's instructions. Forty-eight hours after the first transfection, a second transfection was performed. Cells were maintained at 70–80% confluency and collected 96 h after the first transfection. Knockdown was confirmed by

western blot analysis (list of antibodies and dilutions used are given in Supplementary Table 1).

For studying the effect of *Rbm15*, *Rbm15b*, *Mettl3*, *Ythdf1*, *Ythdf2*, *Ythdf3*, *Ythdc1* and *Ythdc2* knockdown on *XIST*-mediated gene silencing, 20 nM of siRNA targeting each gene were transfected into 100,000 pSM33 ES cells using the Neon transfection system (settings: 1,200 V, 40 ms width, 1 pulse; Invitrogen). At the time of *XIST* induction, the observed knockdown efficiency for all the target genes was greater than 70%. For *Mettl3*, the efficiency was 95%.

Construction of iCLIP libraries. All iCLIP studies were performed on the endogenous proteins. Previous CLIP-based analyses of YTH proteins used overexpressed proteins. Since this can affect the localization and assembly of proteins into multi-protein complexes, we identified antibodies that bound the endogenous proteins for these studies. iCLIP libraries were constructed as described elsewhere with minor modifications³⁰. To improve the efficiency of cell lysis and dissolution of RNA-protein conjugates, cells were lysed in 1% SDS as described previously³¹. In brief, 9×10^6 HEK293T cells were seeded per 10 cm dish 12 h before UV irradiation. Media was discarded and 6 ml of ice-cold PBS was gently added to the cells. Cells were maintained on ice and immediately irradiated once with UV at 254 nm (150 mJ cm^{-2}) in a UV crosslinker (Stratagene 2400). Cells were scraped in PBS using a cell scraper and collected by centrifugation at 200g for 10 min at 4°C. Supernatant was discarded, and cells were gently suspended in 100 μ l of 1% SDS with 10 mM DTT and $10 \times$ protease inhibitors (EDTA-free cComplete mini, Roche) and incubated at 25°C for 10 min to denature the protein complexes. SDS was neutralized with 900 μ l of iCLIP lysis buffer (CLB) without SDS (50 mM Tris-HCl pH 7.4, 100 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate). Lysates were sonicated using a Branson Digital Sonifier Model 450 fitted with 3.125 mm tapered microtip probe on ice at 20% amplitude for 30 s with 2 s ON and 10 s OFF cycle. DNase I and RNase I digestion was performed with 2 μ l of Turbo DNase I (AM2238, Life Technologies) and 10 μ l of different dilutions of RNase I per ml of lysate for 3 min at 37°C. For validation of antibodies for iCLIP and the construction of iCLIP libraries, 1:5 dilution of RNase I (AM2295, Life Technologies) was used as high (H) and 1:150 dilution was used as low (L) concentration RNase. Antibodies were first bound to CLB-washed Protein A/G beads (88803, Thermo Fisher) in CLB (50 mM Tris-HCl pH 7.4, 100 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) followed by incubation at 25°C for 30 min with mixing. Beads were washed twice with CLB.

For validation of antibodies for iCLIP, 500 ng of antibody was used per immunoprecipitation and for the construction of iCLIP libraries, 2–10 μ g of antibody was used. Clarified RNase- and DNase-digested lysates were incubated with antibody bound to Protein A/G-beads at 4°C for 12 h. Further steps of iCLIP library preparation were carried out as described previously³⁰. To avoid cross-contamination of RNA and library PCR products, electrophoresis equipment was treated with 10% commercial bleach for 20 min at 25°C and thoroughly washed with nuclease-free water before use. Replicates were tagged with unique barcodes using the 5' Rtcip primer in reverse transcription. Low-, medium- and high-molecular-mass cDNA libraries were mixed at 1:5:5 molar ratio and sequenced on Illumina HiSeq 2500 from a single end for 50 bases.

Analysis of iCLIP sequence data. Low-quality bases, reads with more than two ambiguous base calls, and adaptor sequences were all removed using FLEXBAR tool (–max-uncalled 2 –min-read-length 15 –pre-trim-phred 20, 3' adaptor: AGATCGGAAGAGCGGTTCAG). Reads were demultiplexed based on 5' barcodes for individual replicates using an in-house Linux shell script. Reads were processed in pooled or separate replicate modes using the CITS analysis pipeline³². In brief, reads were converted to fasta format using fastq_to_fasta tool from FASTX-toolkit and then collapsed to remove PCR amplified duplicates based on sequence using CIMS/fasta2collapse.pl script. The barcode was stripped and added to the name of the read. Reads were aligned to the human genome (hg19) using Novoalign (v3.02.12, NovoCraft Technologies) (Options: –t 85 –l 16 –s 1 –r None). Further analysis until the identification of CITS ($P < 0.0001$) was performed as described previously³². Unique sequence reads that are free of PCR duplicates represent unique RNA-protein binding events. These processed reads are referred to as iCLIP/miCLIP tags (or just tags), and the mapped cluster of processed reads are referred to as tag clusters throughout this study.

Motif enrichment analysis. Analysis of motif enrichment was performed on the sense DNA sequence 20 nucleotides up- and down-stream of the called truncation sites using the MEME suite³³. For this analysis the top 20% of the sites identified as statistically significant (those with $P < 0.0001$), with the highest number of crosslinking induced truncations, were used. Since fewer sites were detected for DC2, all of the sites were used for MEME analysis of DC2-binding sites.

Metagene analysis. Metagenes were constructed for the called CITS/miCLIP-identified m⁶A residues using an in-house Perl annotation pipeline and an R script. In brief, the single-nucleotide sites were mapped to different RNA features (5' UTR, CDS and 3' UTR) of the human genome (hg19). The position of the sites was

normalized to the median feature length of the transcripts to which the sites mapped. A frequency distribution plot was generated by counting the number of sites in contiguous bins on a virtual mRNA transcript, sites whose feature lengths represent the median feature lengths of transcripts under analysis. A Gaussian estimate of kernel density was then plotted as a metagene. For YTH, RBM15 and RBM15B proteins, all statistically significant CITS ($P < 0.0001$) were used and for miCLIP m⁶A, residues identified from poly(A) RNA from ref. 17 were used.

Comparison of iCLIP and miCLIP tag coverage. For comparing iCLIP tags, we calculated normalized tag counts using a previously described approach with minor modifications³⁴. Instead of using read counts per million mapped reads (RPM) normalization to reduce PCR amplification bias, we used unique tag counts obtained from CITS analysis. Each iCLIP tag represents a unique RNA–protein or antibody–m⁶A binding event. The number of unique events from a million such events is proportional within replicates and also comparable across different CLIP libraries. For this, the number of iCLIP tags per million uniquely mapped tags (unique tags per million, uTPM) was calculated at every coordinate on the human genome using the following formula: $uTPM = \frac{t \times 10^6}{T}$ where t = number of unique CLIP tags at a base, T = total number of uniquely mapped unique CLIP tags in the whole CLIP library.

For comparing replicates, the normalized mean tag counts (in uTPM) between replicates at randomly selected ten-thousand and 100-bp bins on the human genome were compared. For comparing various iCLIP/HITS-CLIP/miCLIP data sets, iCLIP data analysed in pooled mode was used. Here, the normalized total tag count in the 10-bp flanking region of 11,530 miCLIP-identified m⁶A residues mapping to mRNA and ncRNA (includes snoRNAs, lncRNA and other ncRNAs) were calculated. Only m⁶A residues in non-BCANN consensus sequence were considered for this analysis. These represent unique sites obtained from merging (mergeBed -s -d 2) of CIMS- and CITS-based m⁶A site calls from ref. 17. All rRNA, tRNA, and mitochondrial genomic miCLIP sites were removed. Tag counting was performed using the bedtools suite. Tag counts (uTPM + 1) were compared using scatter plots and Pearson correlation coefficients (r) were determined in R.

For identification of DC1-preferred m⁶A residues, residuals of simple linear regression model were calculated and sorted in R. The top 1% of sites with highest residuals were selected and annotated. HNRNPA2B1 HITS-CLIP data was obtained from a previously published study³⁵ (GEO accession numbers: GSE70061, SRR2071655 and SRR2071656).

For representation of miCLIP tracks in Figs 3a, 4c and Extended Data Figs 4a, b, 5b, 8d, tag counts from miCLIP data sets using poly(A) RNA and miCLIP data sets using total RNA were added at every genomic position (GEO accession number: GSE63753).

Annotation of CITS. Normalized iCLIP tag-abundance was determined in the 20-bp flanking regions of the RefSeq RNA mapping CITS. Sites were then sorted based on tag abundance, and the top 1,000 sites with the highest normalized tag abundance were annotated using the annotatePeaks.pl script from the Homer package³⁶.

Statistical significance of overlap of RNA-binding sites. To determine the statistical significance of overlap of RBM15 and RBM15B CITS (RBM15, $n = 37$; RBM15B, $n = 56$; $P < 0.0001$ for both) on *XIST*, random sites were generated on the RNA and an overlap with the RBM15 CITS was calculated (± 20 nucleotides) using the bedtools window tool. This was repeated 10,000 times to generate a null distribution for overlap counts. The P -value for the observed overlap between RBM15 and RBM15B was estimated from the null distribution (two-sided). For clusters, random clusters of equal size (median length = 91 nucleotides) were generated on *XIST* and a similar null distribution to CITS was generated. For both comparisons, the same number of random sites or clusters were generated as in RBM15B data set ($n = 30$). Clusters showing a minimum overlap of half-cluster length with the RBM15B clusters were counted. All RBM15 clusters ($n = 30$) overlapped with RBM15B clusters ($n = 30$) on *XIST* ($P < 0.0001$). The RBM15 cluster overlaps with randomly permuted RBM15B clusters, while maintaining the mean cluster size of 91 nucleotides, did not show a similar or greater percentage overlap.

RBM15/15B binding at m⁶A residues. Unique iCLIP tags were aligned to the hg19 genome using STAR aligner (STAR –outSAMtype BAM SortedByCoordinate –outSAMAttributes All –outFilterMultimapNmax 1 –outFilterMismatchNmax 2). For determination of the average RBM15/15B-binding at m⁶A and non-m⁶A sites (both of which are DRACH-consensus sequences), sequence alignment (BAM) files were further processed using deepTools³⁷. Methylated DRACH sites ($n = 14,209$) were obtained by merging miCLIP sites from HEK293 poly(A) and total RNA from ref. 17. A non-methylated DRACH site was identified near each methylated DRACH site within a distance of 20–200 nucleotides in the same transcript in the refseq transcriptome) using an in-house python script. For this purpose, DRACH sites on transcripts with no miCLIP tags were considered to be non-methylated. Heat maps were generated using the plotHeatmap script from the deepTools suite.

Immunoprecipitation of RBM15, RBM15B and METTL3. HEK293T cells were transfected with 10 nM siRNA (Supplementary Table 9) using the Pepmute transfection reagent and then grown to 80% confluency in a 150 mm dish. After 72 h, cells were washed twice with cold PBS, scraped, and collected by centrifugation. The cell pellet was then resuspended in three packed cell volumes of hypotonic buffer (10 mM HEPES pH 7.6, 10 mM KCl, 1 mM EDTA, 0.1 mM EGTA, protease and phosphatase inhibitor cocktail (Pierce)), and incubated on ice for 10 min. Triton X-100 was added to a final concentration of 0.3%, the lysate was briefly vortexed and centrifuged at 15,000g for 1 min at 4°C. Supernatant (cytoplasm) was discarded, and the nuclear pellet was washed with 3 packed cell volumes of hypotonic buffer and centrifuged as before. The pellet was resuspended in 1 ml NP-40 lysis buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% NP-40, protease and phosphatase inhibitor cocktail) and passed through a 21-gauge syringe several times, followed by treatment with 100 U benzonase for 30 min at 37°C. Nuclear lysates were centrifuged at 21,000g for 15 min at 4°C. Immunoprecipitations were carried out with 500 µg of nuclear extracts with 5 µg of antibody at 4°C overnight, followed by a 2 h incubation with 25 µl of Pierce Protein A/G magnetic beads at 4°C. For the co-immunoprecipitation of METTL3–RBM15B, 250 µg of nuclear lysate was used per 5 µg of the METTL3 antibody. Beads were washed five times with NP-40 lysis buffer and proteins were eluted with 1 × Novex Loading buffer with 50 mM dithiothreitol (DTT). The eluent was heat-denatured, electrophoresed, and transferred to a PVDF membrane and probed for different proteins. A list of antibodies and dilutions used for immunoprecipitation and western blot analysis are given in Supplementary Table 1. Quantification of band intensities was performed by the relative quantitation approach using Image Lab software (Bio-Rad, v5.2.1).

RNP immunoprecipitation and quantification of *XIST*. METTL3/DC1/RBM15/RBM15B-bound *XIST* RNA was quantified in the immunoprecipitates obtained from formaldehyde-crosslinked cells using a method previously described³⁸ with some modifications. In brief, siRNA-transfected cells were washed with ice-cold PBS and fixed with 1% formaldehyde in PBS for 10 min at 25°C with gentle rocking. Formaldehyde was quenched by adding glycine to a final concentration of 0.25 M and then incubating at 25°C for 5 min. Fixed cells were washed three times with ice-cold PBS and resuspended in 0.5 ml of radioimmunoprecipitation (RIPA) buffer (50 mM Tris-HCl pH 7.4, 100 mM NaCl, 1% Igepal CA-630, 0.1% SDS, 0.5% sodium deoxycholate) with protease inhibitors (Roche) and 1 mM DTT per 3 million cells. DNA was sheared by sonication on ice twice at 15% amplitude for 2 s ON, 10 s OFF for a total of 30 s. Lysates were incubated on ice for 10 min, and subjected to DNase I and partial RNase I digestion for 3 min at 37°C with mixing (2 µl Turbo DNase I and 5 µl of 1 to 25 times diluted RNase I in PBS per 0.5 ml of lysate). Tubes were immediately transferred to ice and incubated for 5 min. Lysates were then clarified by centrifugation at 21,000g at 4°C for 10 min. Protein (200 µg) was supplemented with SUPERase In RNase inhibitor (100 U ml^{−1}, Thermo Fisher) and then subjected to immunoprecipitation in RIPA buffer. Antibodies targeting METTL3, DC1, RBM15 or RBM15B (2 µg per 10 µl beads; Supplementary Table 1) were first bound to RIPA-buffer-washed Protein A/G magnetic beads (Thermo Fisher). Antibody-bound beads were then washed with RIPA buffer, added to the lysate for immunoprecipitation and incubated at 4°C for 12 h. Rabbit IgG antibody was used as a control. Beads were washed five times with 500 µl RIPA buffer containing 1 M NaCl and 1 M Urea at 25°C and resuspended in 100 µl *eGFP*-RNA (100 pg)-containing RNA elution buffer (50 mM Tris-HCl pH 7.4, 5 mM EDTA, 10 mM DTT, 1% SDS). Formaldehyde-induced crosslinks were reversed by incubation at 70°C for 30 min with mixing. Supernatant was mixed with Trizol LS (Thermo Fisher) and co-immunoprecipitated RNA was purified according to the manufacturer's instructions. Glycoblue (Thermo Fisher) was used to visualize the RNA pellet. Purified RNA was then reverse-transcribed with random hexamers using SuperScript III reverse transcriptase. *XIST* RNA levels were detected by qRT-PCR and normalized to the spike-in *eGFP* RNA levels. Relative *XIST* RNA enrichment was calculated as the ratio of normalized *XIST* RNA levels in protein immunoprecipitation to levels in IgG immunoprecipitates. A very low level of *XIST* RNA was detected in the immunoprecipitate of non-crosslinked cells compared to the crosslinked cells (<1%). Quantification of *XIST* was performed using primer pairs directed against three regions in *XIST*, selected based on the presence of RBM15- and RBM15B-binding sites (see Figs 2b, 3a). These regions were: region 1 (chrX:73,072,444–73,072,560), region 2 (chrX:73,046,651–73,046,776), and region 3 (chrX:73,067,594–73,067,714). Region 1 and 2 contain RBM15/15B-binding sites whereas region 3 lacks RBM15/15B-binding sites. Primers used for quantification are given in Supplementary Table 8. Primer PCR amplification efficiency was between 90 and 100%.

MeRIP qRT-PCR of *XIST* RNA. Total RNA was isolated from HEK293T cells by Trizol extraction according to the manufacturer's instructions and poly(A) RNA was isolated using oligo-d(T)₂₅ magnetic beads (NEB). In total, 5 µg of anti-m⁶A antibody (ab190886, Abcam) was pre-bound to Protein A/G magnetic beads in immunoprecipitation buffer (20 mM Tris-HCl pH 7.5, 140 mM NaCl, 0.05% Triton

X-100) for 2 h. A total of 2.5 µg of poly(A) RNA was mixed with 100 pg of non-m⁶A (*eGFP*, 0.7 kb) and m⁶A-containing spike-in RNAs in 400 µl of immunoprecipitation buffer. Protein A/G beads were then added and incubated at 4 °C for 2 h. Samples were washed five times with immunoprecipitation buffer, and RNA was eluted from the beads by incubating with 400 µl of 0.5 mg ml⁻¹ m⁶ATP for 1 h at 4 °C. Following ethanol precipitation, the input RNA and eluted poly(A) RNA were reverse transcribed with random hexamers and enrichment was determined by qRT-PCR. The spike-in control RNAs were synthesized by *in vitro* transcription. Non-m⁶A RNA (*eGFP*) was transcribed using an *eGFP*-ORF-containing plasmid in the presence of ATP (no m⁶ATP). The m⁶A-containing RNA was transcribed from an artificially synthesized dsDNA template that encoded a 1.6-kb RNA with only one adenosine residue in the presence of m⁶ATP and no ATP.

X-chromosome silencing assay. For this assay, a previously described method³ was used. In brief, siRNA-transfected male or female pSM33 cells were plated on poly-L-lysine or poly-D-lysine (Sigma-Aldrich) and 0.2% gelatin (Sigma-Aldrich)-coated coverslips in wells of a 24-well plate in 2i media. After 48 h, *Xist* RNA expression was induced with doxycycline (2 µg ml⁻¹) (Sigma-Aldrich) in fresh media for 16 h. Control cells received only media. Immediately following incubation, cells were fixed for FISH staining.

For inducing differentiation and induction of *Xist* expression in the female ES cells, 2i media was replaced with MEF media (DMEM, 10% BenchMark FBS; Gemini Bio-products, 1 × L-glutamine, 1 × NEAA, 1 × penicillin and streptomycin; Life Technologies) 12 h after transfection. After another 12 h, cells were treated with 1 µM retinoic acid (Sigma-Aldrich) for 24 h. Untreated cells were maintained in 2i media until fixing.

Cells were then fixed in Histochoice (Sigma-Aldrich) for 10 min, washed with PBS, and subjected to FISH staining and imaging. *Atrx*, *Gpc4*, *Mettl3*, *Rbm15*, *Rbm15b*, *Xist*, *Ythdc1*, *Ythdc2*, *Ythdf1*, *Ythdf2*, and *Ythdf3* RNAs were stained by single-molecule RNA-FISH. They were then imaged and quantified as described in ref. 3. Probe sets and conjugated fluorophores (excitation wavelengths) for FISH probes were TYPE 1-*Xist* (550 nm), TYPE 4-*Gpc4* (488 nm), TYPE 10-*Atrx*, *Rbm15b* (740 nm), and TYPE 6-*Mettl3*, *Rbm15*, *Ythdc1*, *Ythdc2*, *Ythdf1*, *Ythdf2*, and *Ythdf3* (650 nm). Imaging was performed using Nikon Ti Eclipse microscope with the Nikon CFI Plan Apochromat λ DM 60×/1.40 oil objective. Images were processed in Fiji (ImageJ v1.51d)³⁹. To enhance the FISH spot size, Maximum Filter plugin with a radius of 2.0 pixels was applied to the *Gpc4* and/or *Atrx* channels.

DC1- λ N-XIST-(BoxB)₃ RNA tether function assay. For this assay, male mouse pSM33 cells expressing *Xist*-(BoxB)₃ RNA under doxycycline control were used. Cells (1.5 × 10⁵) were co-transfected with 20 nM siMETTL3 or siRBM15/15B and 0.75 µg of pCAG-GW-hYTHDC1- λ N-3×Flag-BSD plasmid using Neon transfection system (10 µl tip, settings: 1,200 V, 40 ms width, 1 pulse) and seeded on coverslips as described for the X-chromosome silencing assay. For the identification of DC1- λ N-3×Flag expressing cells, fixed cells were first subjected to immunofluorescence using mouse anti-Flag antibody (Sigma-Aldrich). Briefly, fixed cells were permeabilized with 0.1% Triton X-100 in PBS at room temperature for 10 min, and blocked with 5% normal goat serum in 0.1% Triton X-100 in PBS at room temperature for 30 min. Cells were then incubated with anti-Flag M2 antibody (Sigma-Aldrich; F3165; dilution 1 to 50) for 1 h at room temperature, followed by washes with 0.1% Triton X-100 in PBS and incubation with secondary antibody (goat anti-mouse IgG antibody-Alexa Fluor 750 conjugate, Thermo Fisher, dilution 1:200) at room temperature for 1 h. The samples were then processed using the RNA-FISH protocol, as described above.

Protein-protein interaction (PPI) network analysis. PINA2 (ref. 28) was used to mine the PPI networks of DC1, its immediate neighbours, the proteins regulating *XIST*-mediated gene silencing (SHARP, HDAC3, HNRNPK, HNRNPU, NCOR2/SMRT, LBR), and components of PRC (polycomb repressor complexes). Protein sub-networks showing interaction with DC1 and an enrichment of transcription repressor gene ontology terms (false discovery rate < 0.05, *P* < 0.05) were curated and filtered for visualization. Networks were imported, visualized, and edited in Cytoscape (v3.3.0)⁴⁰ for image production. To identify potentially novel interactions between DC1 and the proteins contributing to *XIST*-mediated gene silencing, publicly available mass spectrometry data of DC1-associated proteins (PeptideAtlas accession number PASS00835) from ref. 27 was mined. Peptides were first identified by comparing the mass spectrometry spectra with references from the human proteome database (SwissProt) according to ref. 41 (15 p.p.m. peptide mass tolerance and 20 m.m.u. fragment mass tolerance). Identified peptides with natural log(e) scores below -1 and more than two unique peptides were further mined for peptides from proteins known to regulate *XIST*-mediated gene silencing. Identified proteins were manually added to the PPI network.

Determination of relative m⁶A levels by thin layer chromatography. Levels of internal m⁶A in mRNA were determined by thin layer chromatography (TLC) as previously described⁴². In brief, poly(A) RNA (100 ng) was digested with 2 U RNase T1 (Thermo Fisher) for 2 h at 37 °C in the presence of RNasin RNase

Inhibitor (Promega). Five prime ends were subsequently labelled with 10 U T4 PNK (NEB) and 0.4 mBq [γ -³²P]ATP at 37 °C for 30 min followed by removal of the γ -phosphate of ATP by incubation with 10 U apyrase (NEB) at 30 °C for 30 min. After phenol-chloroform extraction and ethanol precipitation, RNA samples were resuspended in 10 µl of water and digested to mononucleotides with 2 U of P1 nuclease (Sigma-Aldrich) for 3 h at 37 °C. Following this, 2 µl of the released 5' monophosphates from this digest were then analysed by 2D-TLC on glass-backed PEI-cellulose plates (Merck-Millipore). The nucleotides were first separated in the first dimension in isobutyric acid with 0.5 M NH₄OH (5:3, v/v), followed by isopropanol, HCl and water at a ratio of 70:15:15 (v/v/v) in the second dimension. Signal acquisition was carried out using a storage phosphor screen (GE Healthcare Life Sciences) at 200 µm pixel size on a Typhoon scanner (GE Healthcare Life Sciences). For quantification, m⁶A was calculated as a percentage of the total of the A, C and U spots, as described previously⁴².

Structured Illumination Microscopy (3D-SIM) and image analysis. HEK293T cells were fixed and subjected to immunofluorescence and single-molecule RNA-FISH staining using a protocol from ref. 43 with some modifications. In brief, siRNA-transfected and non-transfected HEK293T cells were seeded on poly-L-lysine-coated no. 1.5 H (170 µm ± 5 µm) coverslips (poly-L-lysine: 3438-100-01, Trevigen; coverslips: 474030-9000-000, Carl Zeiss) in 6-well plates. After 12–24 h of incubation, cells were washed twice with PBS at 25 °C and fixed with 2% methanol-free formaldehyde (28906, Thermo Fisher) in PBS for 10 min at room temperature. Cells were then washed three times with PBS and permeabilized with permeabilization buffer (1% acetylated BSA (Sigma-Aldrich), 0.3% Triton X-100, 2 mM vanadyl ribonucleoside complexes (NEB) in 1 × PBS) at 25 °C for 60 min. Following permeabilization, cells were incubated with rabbit anti-YTHDC1 antibody (ab122340, Abcam, dilution 1:1,000) in permeabilization buffer for 2 h at 25 °C in a humidified chamber. Cells were then washed with immunofluorescence-wash buffer (0.5% Tween-20 in PBS) three times at room temperature. Each wash was maintained for 5 min on cells with gentle shaking. Cells were further incubated with donkey anti-Rabbit IgG antibody-Alexa Fluor 488 conjugate (A-21206, Thermo Fisher, dilution 1:1,000) for 30 min at 25 °C in a humidified, dark chamber. Following the incubation, cells were washed as before, fixed with 4% formaldehyde in PBS for 10 min at room temperature, and washed with PBS three times. The second formaldehyde fixation immobilizes the primary and secondary antibodies at the target antigen. This step avoids loss of antibodies during the probe-hybridization step of RNA-FISH. Probe hybridization in RNA-FISH uses organic solvent such as formamide that may alter antibody structure thereby affecting its ability to bind the target antigen.

After PBS wash, cells were equilibrated in FISH-wash buffer (10% formamide in 2× SSC buffer diluted from a 20× stock (S6639, Sigma-Aldrich)) for 10 min at room temperature, and then incubated with fluorescently labelled DNA probes against *XIST* (Stellaris FISH probes hXIST w/ Q570, SMF-2038-1, Bioscience Technologies) in Hybridization buffer (10% formamide, 10% dextran sulfate in 2× SSC buffer) at a concentration of 100 nM in a humidified chamber at 37 °C for overnight. Following the incubation, cells were washed twice with FISH-wash buffer at 37 °C for 30 min without shaking. Cells were further washed three times with PBS, and then incubated with DAPI (2 µg ml⁻¹ in PBS) for 15 min at room temperature with gentle shaking. Cells were further washed and maintained in PBS until mounting. Coverslips with fixed and stained cells were mounted in mounting media (Prolong Diamond, P36961, Life Technologies) and quickly sealed with a nail polish. After drying of nail polish, the slides were temporarily stored at 4 °C until imaging.

Cells were imaged by super-resolution 3D-SIM on OMX Blaze 3D-SIM super-resolution microscope (Applied Precision) equipped with a 100×/1.40 numerical aperture UPLSAPO oil objective (Olympus), EMCCD cameras (Photometrics), and 405, 488, 568 nm lasers. Fifteen raw images per plane (5 phases at 3 angles) were captured with a Z-spacing of 0.125 µm using an oil with a refractive index of 1.515. To reduce spherical aberrations, an oil of optimal refractive index was first identified. Image reconstruction and registration was performed using SoftWoRx (GE, v6.5) employing channel-specific optical transfer functions (OTFs) and Wiener filter (settings: 0.0020 for red and green channel, 0.0050 for blue channel). Further processing of 32-bit images was performed using Fiji (ImageJ v1.51d) with in-house JavaScript scripts. Images were converted to 16-bit images. A mask for the *XIST* signal (red) was created on all the slices in Fiji using the thresholding menu option. DC1 (green signal) in the mask was extracted using Fiji's math menu options. 3D Object Counter plugin was then used to count the green objects (DC1 signal) in the *XIST* of the nucleus (*n* = 5, 2 *XIST* and 2 autosomal domains per nucleus). For autosomal domains, areas showing dense DAPI staining were manually selected at the region of interest, DC1 signal (green) was obtained, and 3D objects were counted. Objects here refer to 3D objects identified based on distribution and centre of mass of red or green signal across contiguous image slices. To calculate the percentage fraction of DC1 signal that is localized in the *XIST* territory in various knockdowns, total red (*XIST*) and green (DC1)

objects were also counted in each nucleus separately. Percentage DC1 per *XIST* object was calculated using the following formula:

$$\%DC1 \text{ per } XIST = \frac{n_{gx}}{T_{rx}} \times \frac{100}{T_g}$$

where n_{gx} = number of green objects (DC1) in *XIST* domain, T_{rx} = total number of red (*XIST*) objects, and T_g = total number of green (DC1) objects in the nucleus. Two-tailed Mann–Whitney test was used to calculate statistical significance.

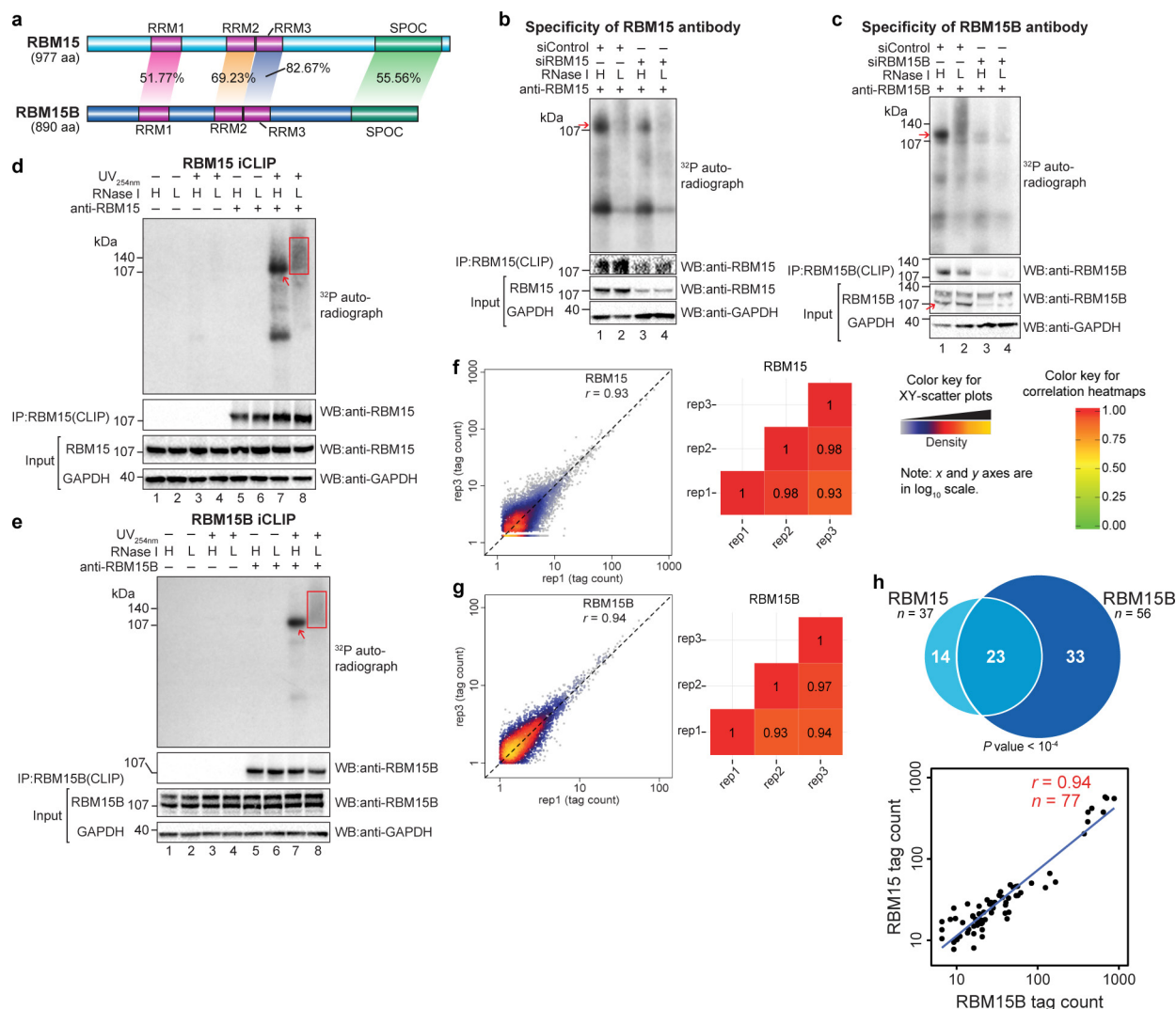
Validation of anti-YTHDC1 antibody for immunofluorescence imaging. HEK293T cells were transfected with pSuperior-EGFP constructs expressing shLacZ or shDC1 shRNA and incubated for 48 h. These cells (20,000 per well) were then seeded on a poly-L-lysine-coated coverslips (coverslips: 1.5 H, 12 mm, round, NC9455457, Fisher Scientific) in 24-well plates. Following a 12-h incubation, cells were processed for immunostaining using the immunofluorescence staining protocol of the 3D-SIM method given above. After the second formaldehyde fixation step, cells were washed three times with $1 \times$ PBS and stained with DAPI, washed, and mounted on slides in mounting media following a method similar to the 3D-SIM method. Slides were stored at 4°C until imaging. DC1 was stained with rabbit anti-YTHDC1 antibody (ab122340, Abcam, 1:1,000) and eGFP (expressed from shRNA expressing plasmid) was stained with chicken anti-GFP antibody (ab13970, Abcam, 1:1,000). Donkey anti-rabbit IgG antibody-Alexa Fluor 568 conjugate (A10042, Thermo Fisher, 1:1,000) and goat anti-chicken IgY antibody-Alexa Fluor 488 conjugate (A-11039, Thermo Fisher, 1:1,000) were used to probe the primary antibodies. Images were captured on a wide-field fluorescence microscope (Nikon Eclipse Ti) using a $60\times$ oil immersion objective. Images were processed on Fiji (ImageJ v1.51d).

Bacterial expression of His6-DF proteins. Full-length DF family cDNA ORFs were PCR amplified from HEK293T oligo-d(T)₂₅-primed cDNA and cloned at NheI and XhoI for DF1 and DF3, NdeI and XhoI for DF2 in pET-28c(+) (Novagen) plasmid. These plasmids were transformed into Rosetta 2(DE3) Singles (Novagen) *Escherichia coli* cells. Bacteria were grown until they reached an OD_{600 nm} of 0.5 and treated with 0.1 mM IPTG at 18°C for 1–4 h to allow a comparable level of protein expression. Time points showing a similar level of protein expression for all the DF proteins were only analysed by western blot. DNA oligonucleotides used for amplification of the cDNA ORFs are given in Supplementary Table 8.

Enrichment of DC1-binding RNA motifs in different RNA and genomic features. For this analysis, all the 35,823 CIT sites were used. CITs were first mapped to the different genomic and RNA features in the hg19 genome using the annotation script, annotatePeaks.pl, from the Homer package. Sites mapping to rRNA, tRNA and the mitochondrial genome were discarded. For every site, strand-specific DNA sequence (± 20 nucleotides) was obtained from the hg19 genome. An enrichment of DC1-binding RNA motifs (DRACH, MTTAH, and KTAHC) in different RNA/genomic features was determined using Centrimo tool in the MEME suite.

Confirmation of X-chromosome silencing by RT-qPCR. Total RNA was extracted and purified from 1×10^6 siRNA-transfected pSM33 cells using RNeasy Mini Kit (Qiagen) and DNA was removed by digestion with RNase-free DNase Set (Qiagen). DNA-free RNA (500 ng) was used to make cDNA with random hexamer using SuperScript III reverse transcriptase (Invitrogen) following the manufacturer's instructions. Relative expression of genes *Gpc4* and *Atrx* relative to *Gapdh* was quantified by qPCR using the LightCycler 480 SYBR Green I Master Mix (Roche). Primer information is given in Supplementary Table 8.

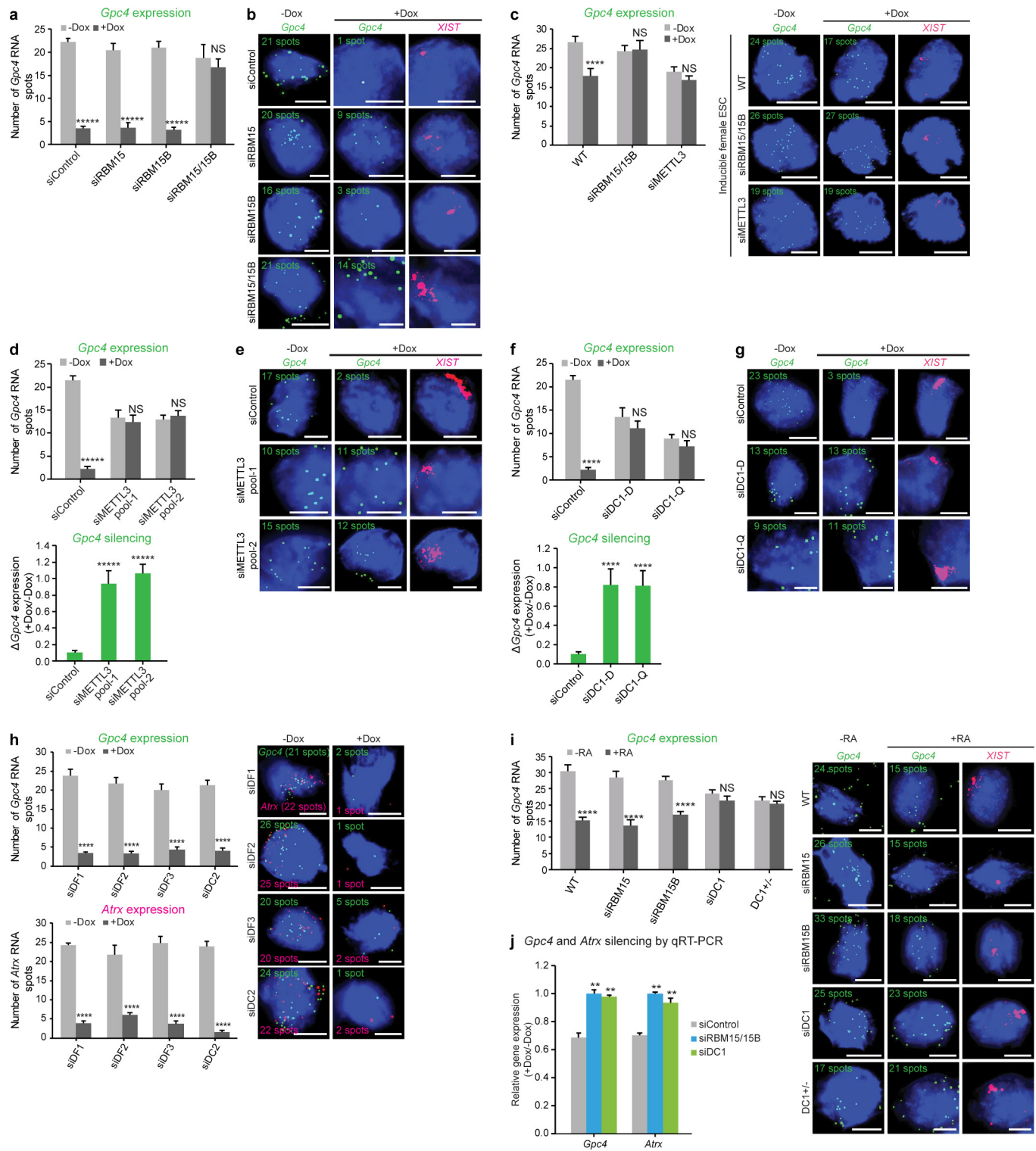
29. Engreitz, J. M. *et al.* The *Xist* lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).
30. Huppertz, I. *et al.* iCLIP: protein-RNA interactions at nucleotide resolution. *Methods* **65**, 274–287 (2014).
31. Meyer, K. D. *et al.* 5' UTR m(6)A promotes cap-independent translation. *Cell* **163**, 999–1010 (2015).
32. Weyn-Vanhenryck, S. M. *et al.* HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. *Cell Reports* **6**, 1139–1152 (2014).
33. Ma, W., Noble, W. S. & Bailey, T. L. Motif-based analysis of large nucleotide data sets using MEME-ChIP. *Nat. Protocols* **9**, 1428–1450 (2014).
34. Singh, G. *et al.* The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. *Cell* **151**, 750–764 (2012).
35. Alarcón, C. R. *et al.* HNRNPA2B1 is a mediator of m(6)A-dependent nuclear RNA Processing events. *Cell* **162**, 1299–1308 (2015).
36. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
37. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–91 (2014).
38. Niranjanakumari, S., Lasda, E., Brazas, R. & Garcia-Blanco, M. A. Reversible cross-linking combined with immunoprecipitation to study RNA-protein interactions in vivo. *Methods* **26**, 182–190 (2002).
39. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
40. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
41. Link, A. J. & LaBaer, J. in *Proteomics: A Cold Spring Harbor Laboratory Course Manual*. Ch. **8**, 117–142 (Cold Spring Harbor, 2009).
42. Jia, G. *et al.* N⁶-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **7**, 885–887 (2011).
43. Kochan, J., Wawro, M. & Kasza, A. Simultaneous detection of mRNA and protein in single cells using immunofluorescence-combined single-molecule RNA-FISH. *Biotechniques* **59**, 209–212, 214, 216 passim (2015).
44. Uranishi, H. *et al.* The RNA-binding motif protein 15B (RBM15B/OTT3) acts as cofactor of the nuclear export receptor NXF1. *J. Biol. Chem.* **284**, 26106–26116 (2009).
45. Geula, S. *et al.* Stem cells. m⁶A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science* **347**, 1002–1006 (2015).
46. Liu, N. *et al.* N⁶-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* **518**, 560–564 (2015).
47. Wei, C. M., Gershowitz, A. & Moss, B. 5'-Terminal and internal methylated nucleotide sequences in HeLa cell mRNA. *Biochemistry* **15**, 397–401 (1976).
48. Bylund, L., Kytölä, S., Lui, W. O., Larsson, C. & Weber, G. Analysis of the cytogenetic stability of the human embryonal kidney cell line 293 by cytogenetic and STR profiling approaches. *Cytogenet. Genome Res.* **106**, 28–32 (2004).
49. Lin, Y. C. *et al.* Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat. Commun.* **5**, 4767 (2014).
50. Cao, Q. *et al.* The central role of EED in the orchestration of polycomb group complexes. *Nat. Commun.* **5**, 3127 (2014).
51. Wilkinson, F. L. *et al.* Emerin interacts in vitro with the splicing-associated factor, YT521-B. *Eur. J. Biochem.* **270**, 2459–2466 (2003).
52. Hartmann, A. M., Nayler, O., Schwaiger, F. W., Obermeier, A. & Stamm, S. The interaction and colocalization of Sam68 with the splicing-associated factor YT521-B in nuclear dots is regulated by the Src family kinase p59(fyn). *Mol. Biol. Cell* **10**, 3909–3926 (1999).



Extended Data Figure 1 | Validation of RBM15 and RBM15B antibodies for iCLIP, construction and comparison of iCLIP library replicates.

a, RBM15 and RBM15B exhibit high sequence homology. RBM15 and RBM15B comprise three RRM domains (RRM1, 2 and 3, all in purple) and a C-terminal SPOC domain (green). These domains show high sequence identity between RBM15 and RBM15B (indicated on the shaded areas that connect the compared regions). RRM, RNA recognition motif; SPOC, Spen paralogue and orthologue C-terminal. **b**, **c**, Validation of specificity of RBM15 and RBM15B antibodies for iCLIP, performed using immunoprecipitation. In each experiment, we used high (H) and low (L) RNase, as per the iCLIP validation protocol³⁰ (see Methods). The bottom western blots are loading control (GAPDH). To confirm knockdown, RBM15 and RBM15B protein levels are shown. Additionally, we show the amount of protein in the anti-RBM15 or anti-RBM15B pull-downs. These experiments confirm that the RBM15 and RBM15B are knocked down after siRNA transfection. **d**, **e**, Autoradiograms of the samples used for the RBM15 and RBM15B iCLIP experiments. Shown are the representative autoradiograms from the nitrocellulose blots of samples used for preparing the RBM15 and RBM15B iCLIP library. The excised portion of the membrane is shown (red square). The red arrow indicates the position of RBM15 and RBM15B protein after high RNase treatment that matches with the size seen in **b** and **c** respectively. Both RBM15 and RBM15B show specific RNA–protein conjugates of expected size with a minimal contamination of RNA–protein conjugates of other sizes. **f**, **g**, RBM15 and RBM15B iCLIP replicates show reproducible iCLIP tag coverage on the

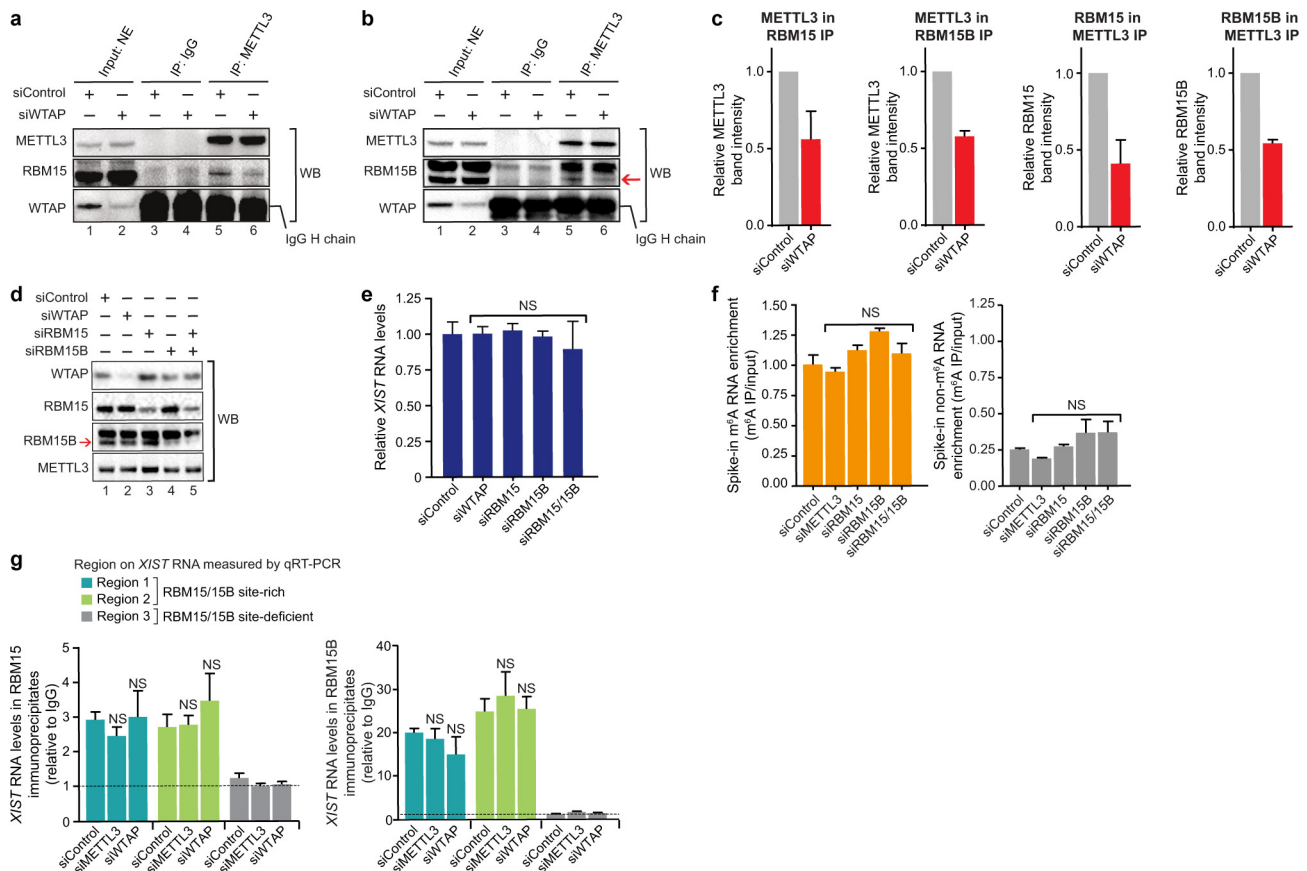
human genome. Three iCLIP library replicates were prepared for RBM15 and RBM15B. We compared the normalized tag counts of replicates in 100 nucleotide bins in the human genome on scatter plots, and estimated the Pearson correlation coefficient (r). Shown are the representative scatter plots (left), and heat maps (right) showing the obtained r value in multiple pairwise replicate comparisons. rep1–rep3, replicate 1–replicate 3 for each protein; RBM15 in **f** and RBM15B in **g**. The x and y axes of the scatter plots represent normalized tag counts in uTPM in 100 nucleotide bins on the human genome in rep1 and rep3, respectively. Correlation values are indicated on each tile. From this analysis, RBM15 and RBM15B iCLIP replicates show a similar, highly reproducible iCLIP tag coverage on the human genome. The diagonal dashed line in scatter plots represents reference trend line for a perfect correlation ($r = 1$, $x = y$). **h**, RBM15 and RBM15B show similar binding preferences on *XIST*. Each of the 30 clusters in the RBM15 data set overlapped with the clusters in the RBM15B data set. We also examined the CITS induced by RBM15 and RBM15B. CITS are single-nucleotide sites that represent direct contacts of these proteins with *XIST* (Supplementary Tables 3, 4). Most RBM15 CITS (23 out of 37) overlapped with RBM15B CITS (top). This overlap was statistically significant ($P < 0.0001$) based on a permutation analysis in which we measured the overlap of randomly selected sites on *XIST* for RBM15 and RBM15B (see Methods). Lastly, a pairwise analysis of iCLIP tag density at each CITS showed that RBM15 and RBM15B binding was highly correlated (bottom).



Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | Quantification of X-linked gene silencing upon knockdown of m⁶A readers and writers. **a, b**, Quantification of *Gpc4* spots upon *Rbm15* and *Rbm15b* knockdown (Fig. 1b, c). The number of *Gpc4* spots before and after *XIST* induction (–Dox and +Dox, respectively) (**a**). Representative RNA-FISH images with DAPI-stained nuclei with *Gpc4* spots (green) and *XIST* staining (pink, last column) are shown (**b**). The number of *Gpc4* spots is indicated on each FISH image. Scale bar, 5 μ m. Data in **a** are mean \pm s.e.m. NS, not significant; **** $P < 0.0001$ relative to Dox-deficient control by unpaired two-sample *t*-test. **c**, m⁶A modification is necessary for *XIST*-mediated gene silencing in female pSM33 cells. Quantification of *Gpc4* RNA spots with and without induction of *XIST* expression (left). Representative RNA-FISH images showing *Gpc4* RNA spots (green) with DAPI-stained nuclei (right). Wild-type (WT) cells show a normal *XIST*-induced silencing whereas *Gpc4* spots are partially reduced (24 to 17 spots). Similar to male ES pSM33 cells, female ES cells fail to show *XIST*-mediated gene silencing upon knockdown of *Rbm15/15b* or *Mettl3*. Error bars mean \pm s.e.m. for 50 cells per sample. NS, not significant; **** $P < 0.0001$, relative to no-doxycycline control by unpaired two-sample *t*-test. **d, e**, Similar to Fig 3c, d, shown is an siRNA pool that targets a (different) region on *Mettl3*. The data from Fig. 3c, d for the siRNA pool 1 is also shown here for comparison. In both the siControl and siMETTL3-transfected cells, *XIST* shows aggregation consistent with its interaction with the X chromosome. Thus, early steps of *XIST* interaction with the X chromosome may not require m⁶A. *Gpc4* counts (**d**, top) and the change in transcription, as measured by the ratio of *Gpc4* +Dox/–Dox. Notably, there is a reduction in *Gpc4* and *Atrx* spots (see Fig. 3d) in siMETTL3-transfected cells, even in the absence of *XIST* expression. Representative FISH images with DAPI nuclear stain in blue, *Gpc4* in green and *XIST* in pink (**e**). Following Dox treatment, the number of *Gpc4* spots is markedly reduced in the control-transfected cells. However, after knockdown of *Mettl3*, the number of *Gpc4* mRNA spots remain unchanged. Scale bars, 5 μ m. Data in **d** are mean \pm s.e.m. across 50 cells. NS, not significant; **** $P < 0.0001$ relative to no-doxycycline control (top graph) and siControl (bottom graph) by unpaired two-sample *t*-test. **f, g**, Similar to **d** and **e**, we show a defect in *XIST*-mediated silencing upon silencing of *Ythdc1* as shown in Fig. 4d, e using multiple siRNA pools from different vendors. Targeting a different region of DC1 using a siRNA pool (siDC1-Q) prevents *XIST*-mediated

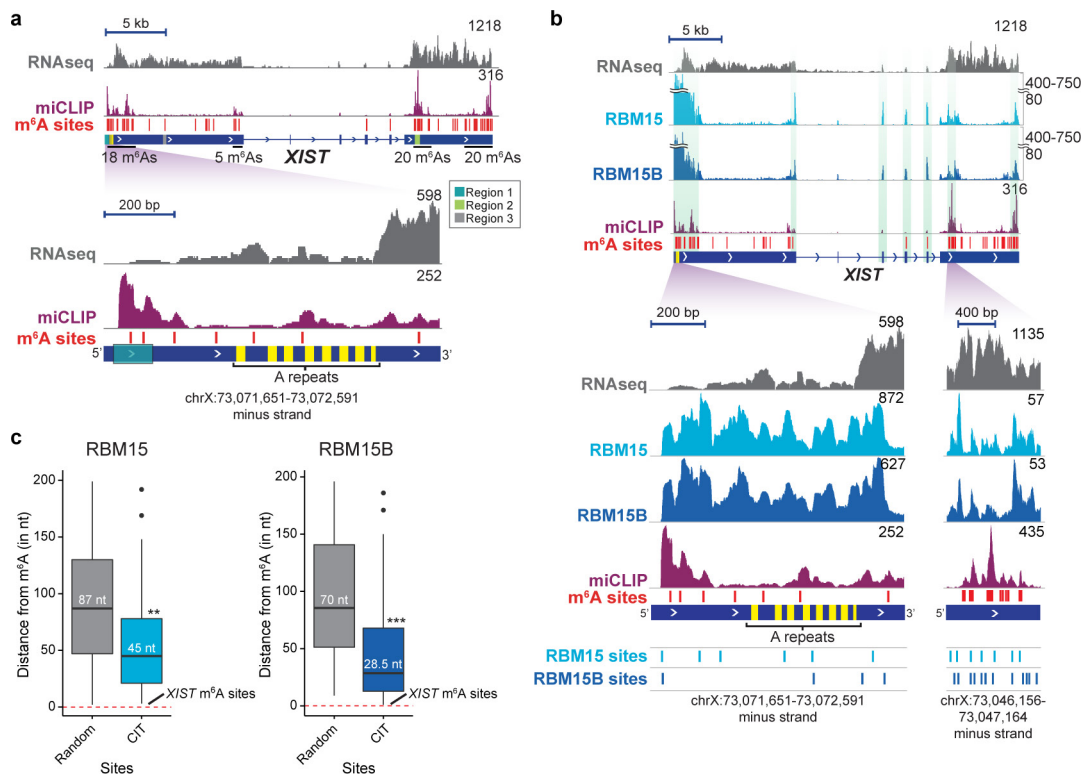
gene silencing. The data from Fig. 4d, e for the Dharmacon siRNA pool is shown alongside. Data in **f** are mean \pm s.e.m. across 50 cells. NS, not significant; **** $P < 0.005$ relative to no-doxycycline control (top graph) and siControl (bottom graph) by unpaired two-sample *t*-test. **h**, DF1, DF2, DF3 and DC2 do not mediate *XIST*-mediate gene silencing. Quantification of *Gpc4* (top left) and *Atrx* (bottom left) RNA-FISH spots is shown. Representative FISH images with DAPI-stained nuclei (blue) with *Gpc4* (green) and *Atrx* (red) spots are shown (right). The number of detected RNA spots for both the genes are indicated on each FISH image. Scale bars, 5 μ m. Data are mean \pm s.e.m. across 50 cells from one experiment. **** $P < 0.0001$ relative to control (–Dox) by unpaired two-sample *t*-test. **i**, RBM15/15B and DC1 mediate *XIST*-mediate gene silencing in differentiating wild-type female ES cells. Quantification of *Gpc4* RNA expression was performed in female mouse ES cells in response to retinoic acid-induced (+RA) differentiation by RNA-FISH (left). Representative FISH images showing DAPI-stained nuclei (blue), *Gpc4* RNA (green), and *XIST* (pink) are shown (right). Wild-type cells exhibit normal *Gpc4* silencing in response to retinoic acid treatment. Single knockdown of either *Rbm15* or *Rbm15b* also exhibited normal silencing of *Gpc4*. Double knockdown resulted in no *XIST* expression (C.-K.C. and M.G., data not shown), reminiscent of the lack of *XIST* expression in METTL3-deficient ES cells⁴⁵. CRISPR-mediated homozygous knockout of *DC1* (*Ythdc1*^{–/–}) cells could not be recovered, suggesting that deletion of this gene is lethal. However, heterozygous knockout of *DC1* (*Ythdc1*^{–/+}) impaired *Gpc4* silencing in response to retinoic acid in these cells. These data support the idea that DC1 is required for silencing of X-linked genes during ES cell differentiation. **** $P < 0.0001$ relative to control by unpaired two-sample *t*-test. **j**, qRT-PCR-based validation of effects of RBM15/15B and DC1 on *XIST*-mediated gene silencing. Gene expression level after *XIST* induction (+Dox) was normalized to Gapdh before *XIST* induction (–Dox) in both the siControl and siRbm15/siRbm15b double-knockdown sample. Quantification of the change in gene transcript levels upon expression of *XIST* is shown for *Gpc4* and *Atrx*. Dox-induced *XIST* expression led to reduced transcription of both the genes in Control knockdown cells. However, *Rbm15* and *Rbm15b* double knockdown and *DC1* knockdown failed to show *XIST*-induced silencing. ** $P < 0.01$ relative to siControl-transfected cells by unpaired two-sample *t*-test.



Extended Data Figure 3 | Reciprocal co-immunoprecipitation of METTL3-RBM15/15B complex, validation of WTAP, RBM15, and RBM15B knockdown and their lack of effect on XIST levels.

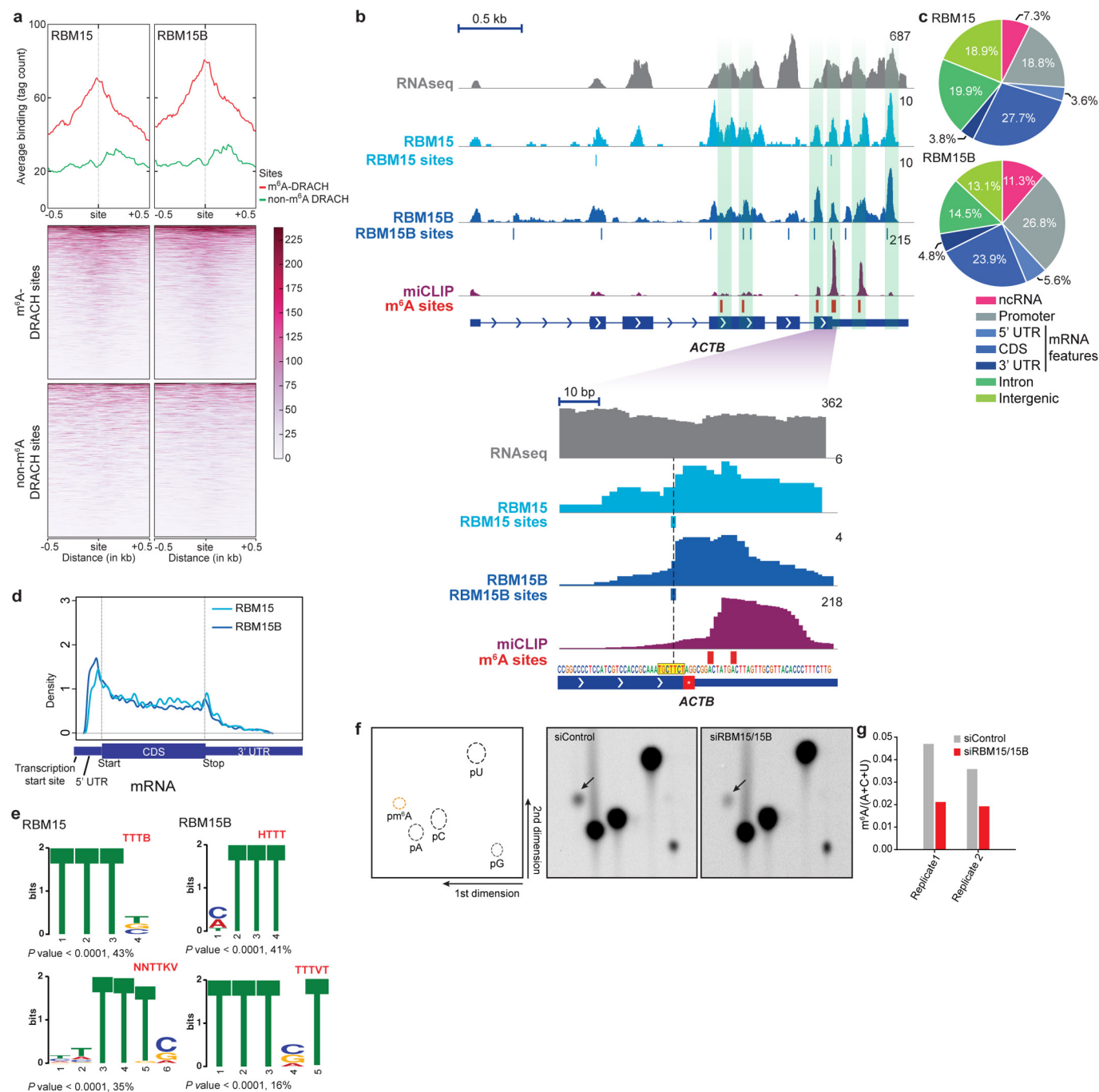
a, b, Confirmation of WTAP-dependent METTL3-RBM15/15B interaction by reciprocal co-immunoprecipitation. METTL3 was immunoprecipitated using an antibody against the endogenous protein from nuclear extracts of the siControl- and siWTAP-transfected HEK293T cells under native conditions. Both RBM15 and RBM15B were detected in the METTL3 immunoprecipitates by western blot. The binding of both these proteins was significantly reduced in siWTAP-transfected cells, indicating that METTL3 interacts with RBM15/15B in a WTAP-dependent manner to form a RBM15/15B-WTAP-METTL3 complex. IgG heavy chain signal prevents visualization of WTAP; however, knockdown is seen in the input sample. **c**, Relative protein band intensities for RBM15/15B-METTL3 and reciprocal co-immunoprecipitation experiments shown in Fig. 2a and Extended Data Fig. 3a, b, respectively. For METTL3 in RBM15 IP, $n = 3$; METTL3 in RBM15B IP, $n = 3$; RBM15 in METTL3 IP, $n = 7$; and RBM15B in METTL3, $n = 3$. **d**, Confirmation of WTAP, RBM15, and RBM15B knockdown. siRNA-transfected HEK293T cell lysates used for assays in Figs 2b, 3b were probed for protein levels using western blot analysis. Knockdown resulted in a significant reduction in the corresponding proteins. None of the siRNAs affect METTL3 levels.

The antibody for RBM15B recognizes a doublet, but only the lower band is lost after the knockdown. The specificity of this antibody for iCLIP is demonstrated in Extended Data Fig. 1c, e. **e**, Knockdown of WTAP, RBM15 and RBM15B, as well as double knockdown of RBM15 and RBM15B do not affect XIST RNA levels. Quantification of XIST levels by qRT-PCR from RNA purified from siRNA-transfected cells shows no significant change in XIST RNA levels. **f**, Validation of the anti-m⁶A antibody approach for pulldown of methylated XIST RNA. To validate the XIST quantification used in Fig. 3b, we used a control spike-in RNAs with a single m⁶A, and an eGFP control RNA with no m⁶A residues. Unlike the m⁶A RNA (left), the non-methylated RNA (right) is de-enriched in the immunoprecipitation sample. **g**, RBM15/15B bind XIST in m⁶A-independent manner. RBM15/15B binding of XIST in cells deficient in components of the m⁶A methylation machinery (METTL3 and WTAP) is shown. RBM15 and RBM15B were immunoprecipitated and XIST levels were determined by qRT-PCR at three regions (regions 1–3 refer to Fig. 2b, 3a and Extended Data Fig. 4a). XIST binding to RBM15 and RBM15B remains unchanged upon METTL3 and WTAP knockdown at region 1 and 2 where RBM15/15B both show binding. Thus, RBM15 and RBM15B are not binding to XIST in an m⁶A-dependent manner and are not m⁶A readers. At region 3, where both proteins do not show any binding, a basal level of amplification was seen similar to the level detected in IgG control. NS, not significant relative to siControl transfected cells by unpaired two-sample *t*-test (**e–g**).



Extended Data Figure 4 | Zoomed-in views of miCLIP, RBM15 and RBM15B iCLIP tracks on *XIST*. **a**, m⁶A residues are broadly distributed along *XIST*. Shown are m⁶A residues mapped in *XIST* using miCLIP¹⁷; these sites are indicated with red lines. Total RNA at every genomic position are shown in purple. RNA-seq read distribution is shown in grey. Many of the m⁶A sites are clustered in a 2 kb domain surrounding the A-repeat (yellow) region. The zoomed-in region shows m⁶A sites (red lines) and miCLIP tag distribution in a 1-kb region closest to the A-repeat region. Region 1, which contains RBM15/15B-binding sites (see Fig. 2b) is also indicated. **b**, **c**, RBM15 and RBM15B bind *XIST* near m⁶A sites. To determine whether RBM15/15B-binding sites are in proximity to known m⁶A sites, we compared the iCLIP tag clusters with m⁶A sites on *XIST*. Shown in **b** are the RBM15 and RBM15B iCLIP, and miCLIP tag distributions on *XIST*. m⁶A sites are marked with red bars above the *XIST* gene model. Vertical green shaded boxes mark the regions

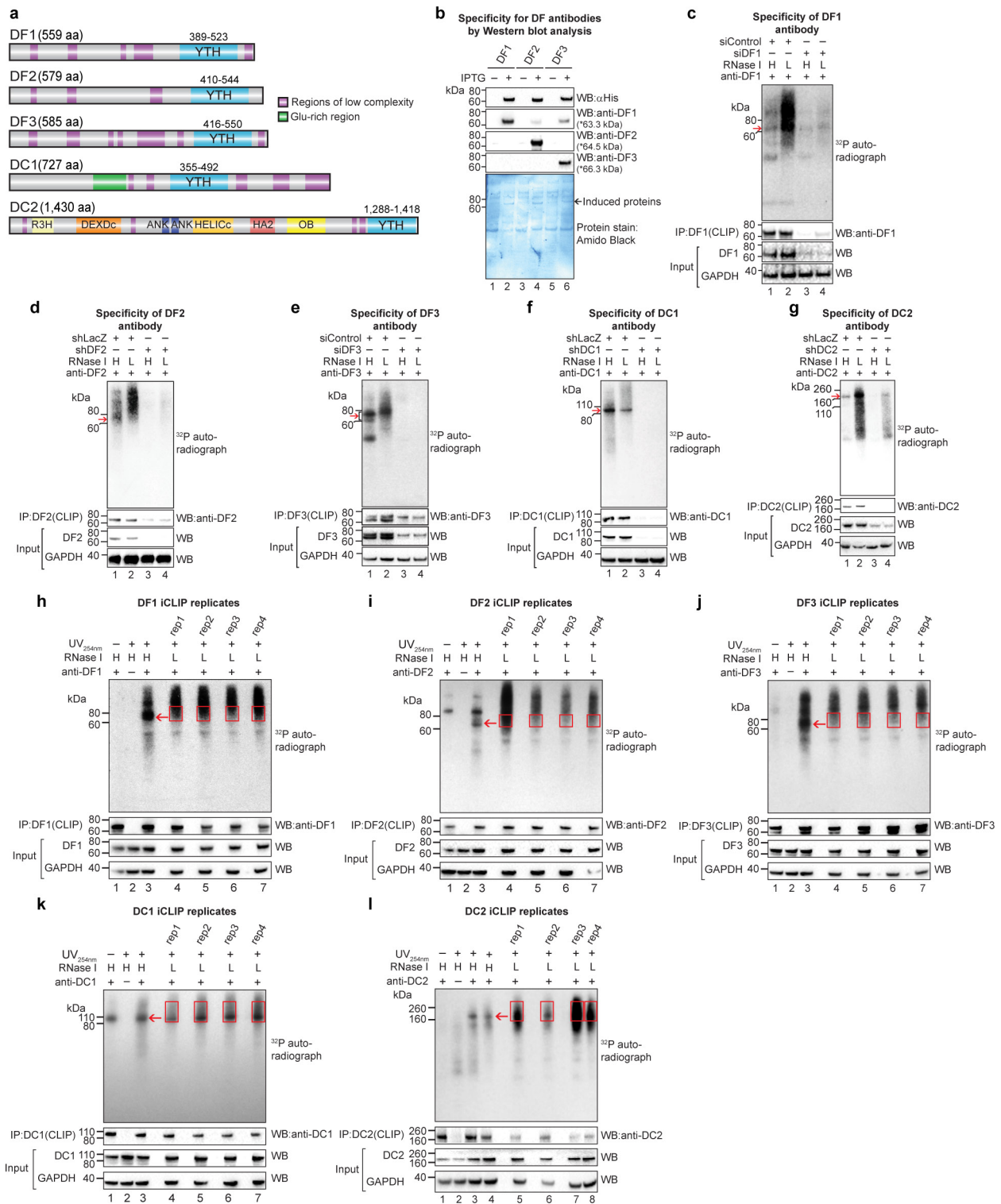
of miCLIP and RBM15/15B iCLIP tag cluster alignments. A zoomed-in view of a region with high-tag abundance (bottom left) and another with low-tag abundance (bottom right), show examples of m⁶A sites that are in proximity to RBM15B and RBM15B tag clusters. Normalized tags are shown in uTPM. In **c**, the median distance of RBM15 (left) and RBM15B (right) CITS to the nearest m⁶A site on *XIST* was determined and compared with a randomly permuted data set of RBM15- and RBM15B-binding sites. RBM15/15B-binding sites show a marked proximity to m⁶A compared to randomly positioned RBM15/15B sites (RBM15, ***P* = 0.0026, number of permutations, 10,000; RBM15B, ****P* = 0.0001, number of permutations, 10,000). This proximity is not due to RBM15 or RBM15B itself binding m⁶A as its binding to *XIST* was unaffected by *METTL3* or *WTAP* knockdown (Extended Data Fig. 3g). The red dashed line indicates the location of m⁶A sites.



Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | RBM15 and RBM15B bind near m⁶A sites on mRNA. **a**, RBM15/15B binds at-or-near-to m⁶A sites throughout the transcriptome, including at m⁶A sites in *XIST* and *ACTB* mRNA. Shown are plots with an average binding-per-base around m⁶A (red curve) or non-m⁶A DRACH (green curve) sites for RBM15 (top left) and RBM15B (top right). The bottom two panels present the tag count per base around m⁶A or non-m⁶A DRACH sites as heat maps. Each row in the heat map is an m⁶A or non-m⁶A site. RBM15 and RBM15B show increased binding at or near m⁶A sites than at non-methylated DRACH sites (~3–4-fold higher). **b**, RBM15 and RBM15B bind near m⁶A sites on mRNA. Shown is the RNA-seq read (grey), and iCLIP (light blue, RBM15; dark blue, RBM15B) and miCLIP (purple) tag distribution on *ACTB* mRNA. iCLIP CITS sites are indicated below their respective tracks. miCLIP-identified m⁶A sites are indicated with red bars. Both proteins (light versus dark blue tracks) show a similar binding profile on *ACTB* mRNA, with considerable overlap of miCLIP tags at various regions along the sequence (vertical green shading). A zoomed-in view of the tag distribution is shown in the bottom panel. The sense DNA sequence of the zoomed-in region is shown above the gene model. A vertical dotted black line running through the middle of the tracks connects the RBM15/15B-binding sites with the DNA sequence that indicates the sequence at the binding site (highlighted yellow). At single-nucleotide resolution RBM15/15B binds a U-rich sequence near m⁶A sites on mRNA also. The binding sites show a clear separation (5 nucleotides) from the putative m⁶A-containing GAC sequence (red bars). RNA-seq reads are shown in absolute read counts, iCLIP and miCLIP tags are shown in uTPM. **c, d**, Genomic and transcriptomic distribution of RBM15- and RBM15B-RNA-binding sites. To determine the types of RNA sequence that contain bound RBM15 and RBM15B, the top 1,000 iCLIP CITS ($P < 0.0001$) with the highest iCLIP tag coverage (in uTPM) were mapped to different features of the human genome and the overall distribution was determined. Sites mapped to mRNA (blue) represent roughly an equal fraction of all the binding sites

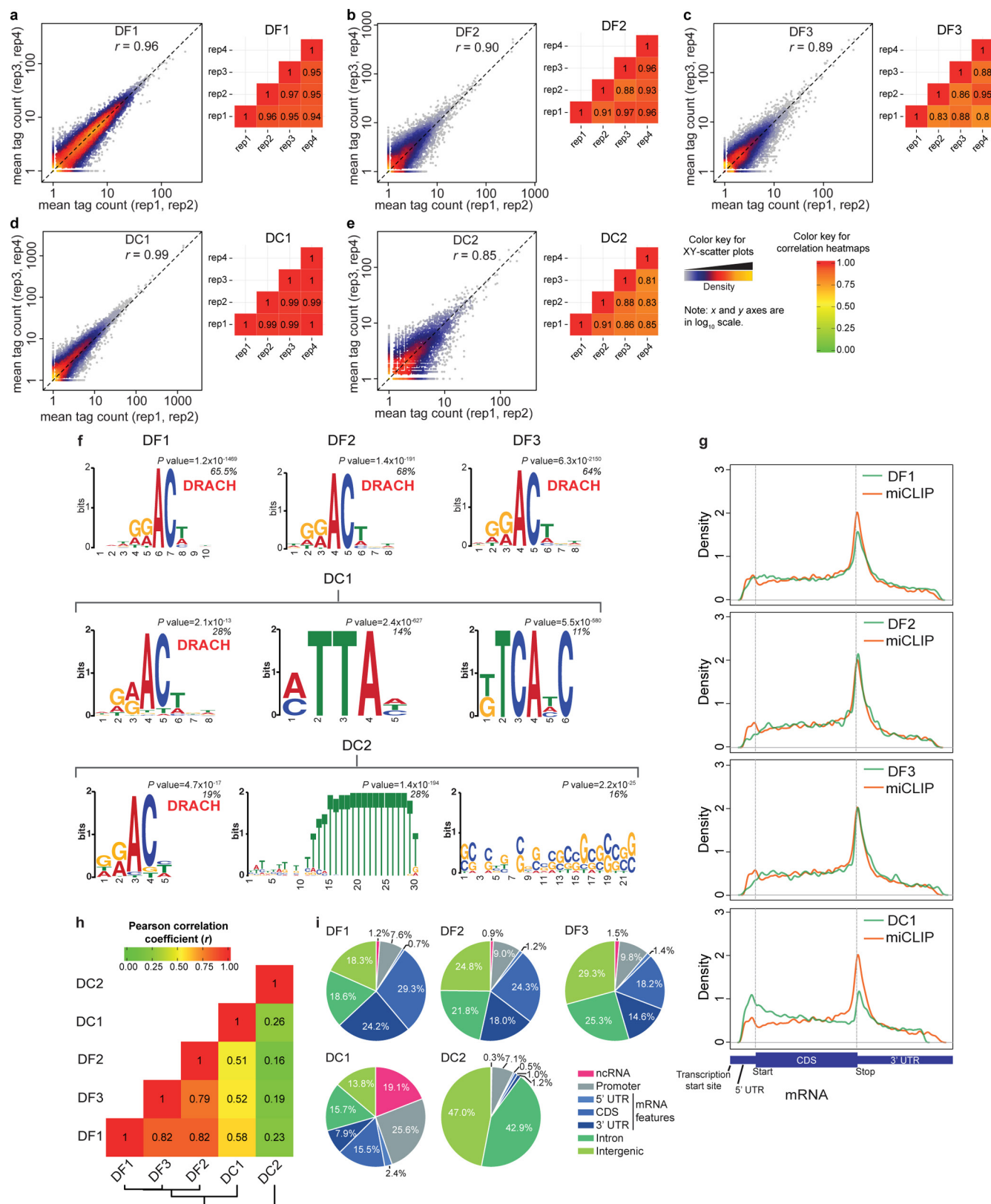
of the proteins (~35%). To determine the overall distribution of the RNA-binding sites in mRNA, we further plotted the distribution of all the RBM15- and RBM15B-binding sites on a virtual transcript (shown in **d**). Metagenes for both RBM15- and RBM15B-binding sites show a similar distribution of the binding sites on the different features of mRNA. Although this metagene shows coverage all along mRNA, as is seen with m⁶A, this distribution does not match the m⁶A metagene. CDS, coding sequence; UTR, untranslated region. **e**, RBM15 and RBM15B bind U-rich RNA consensus motif. Shown are motifs enriched in both RBM15- and RBM15B-binding sites and the percentage distribution of the sites containing the identified motif is indicated below each motif. U-rich RNA-binding motifs (shown as T in this genome-based alignment) were significantly enriched in the sequence at or around the iCLIP-identified RBM15- and RBM15B-binding sites ($P < 0.0001$). The absence of an m⁶A-like DRACH motif for both the proteins indicates that RBM15/15B does not directly bind m⁶A or DRACH sequences. Notably, the U-rich motif seen with RBM15/15B resembles the uracil-rich HNRNPC-binding motif, which may account for the previously observed proximity between m⁶A and HNRNPC-binding sites⁴⁶. **f, g**, Knockdown of *RBM15B* and *RBM15B* reduced m⁶A levels in cellular mRNA. Schematic diagram of a 2D-TLC (left, **f**) showing the migration pattern of monophosphate nucleotides after TLC separation. Shown are relative positions of m⁶A (orange dotted circle) and those of adenosine (A), cytosine (C), and uracil (U) (black dotted circles). Arrows indicate the direction of solvent migration in the two dimensions. Middle and right panels show radiochromatograms obtained from 2D-TLC of poly(A) RNA from control and *RBM15/RBM15B* double-knockdown HEK293T cells. Double knockdown of *RBM15* and *RBM15B* leads to a considerable decrease in m⁶A levels in mRNA (spots marked with black arrow in the middle and right panel). Quantification of m⁶A levels calculated using m⁶A:A + C + U ratio from mononucleotide intensity in two independent biological replicates (**g**).



Extended Data Figure 6 | See next page for caption.

Extended Data Figure 6 | Validation experiments for iCLIP of YTH proteins, anti-YTH antibodies, and library construction. **a**, Schematic representation of domain structures of human YTH proteins: DF1, DF2, DF3, DC1 and DC2. The YTH domain (blue) is located internally in DC1, while it is at the C-terminal region in the other proteins. DC1 has a different domain organization to DC2 and the similar DF proteins. The low-complexity and Glu-rich regions are indicated, as are the R3H, DEXDc, ankyrin repeats (ANK), HELICc, HA2 and OB-fold domains. The length of the protein is indicated next to each protein name. **b**, Validation of DF1, DF2 and DF3 antibody specificity via western blot. Full-length DF1, DF2, and DF3 were expressed as His6-fusion proteins in *E. coli*. IPTG was used as an inducer of protein expression (–, non-IPTG-treated; +, IPTG-treated). For anti-DF1, His6-DF1 was the major band detected but trace levels of His6-DF2 and His6-DF3 could be detected at longer exposure times. Thus, anti-DF2 and anti-DF3 antibodies are highly specific, while anti-DF1 shows a strong preference towards DF1 over the other DF proteins. **c–g**, Confirmation of iCLIP antibody pulldown specificity. Autoradiograms of the ^{32}P -labelled RNA-crosslinked protein conjugates on nitrocellulose membrane (top) for DF1 (**c**), DF2 (**d**), DF3 (**e**), DC1 (**f**) and DC2 (**g**) are shown. High (H) and low (L) RNase are used in accordance with the iCLIP validation protocol³⁰ (see Methods). The red arrow indicates the expected size of the YTH protein. In each case, knockdown of the YTH protein mRNA (lanes 3 and 4) abolished RNA pulldown. GAPDH was used as a loading control. To confirm knockdown, protein levels in the input samples and in the anti-YTH

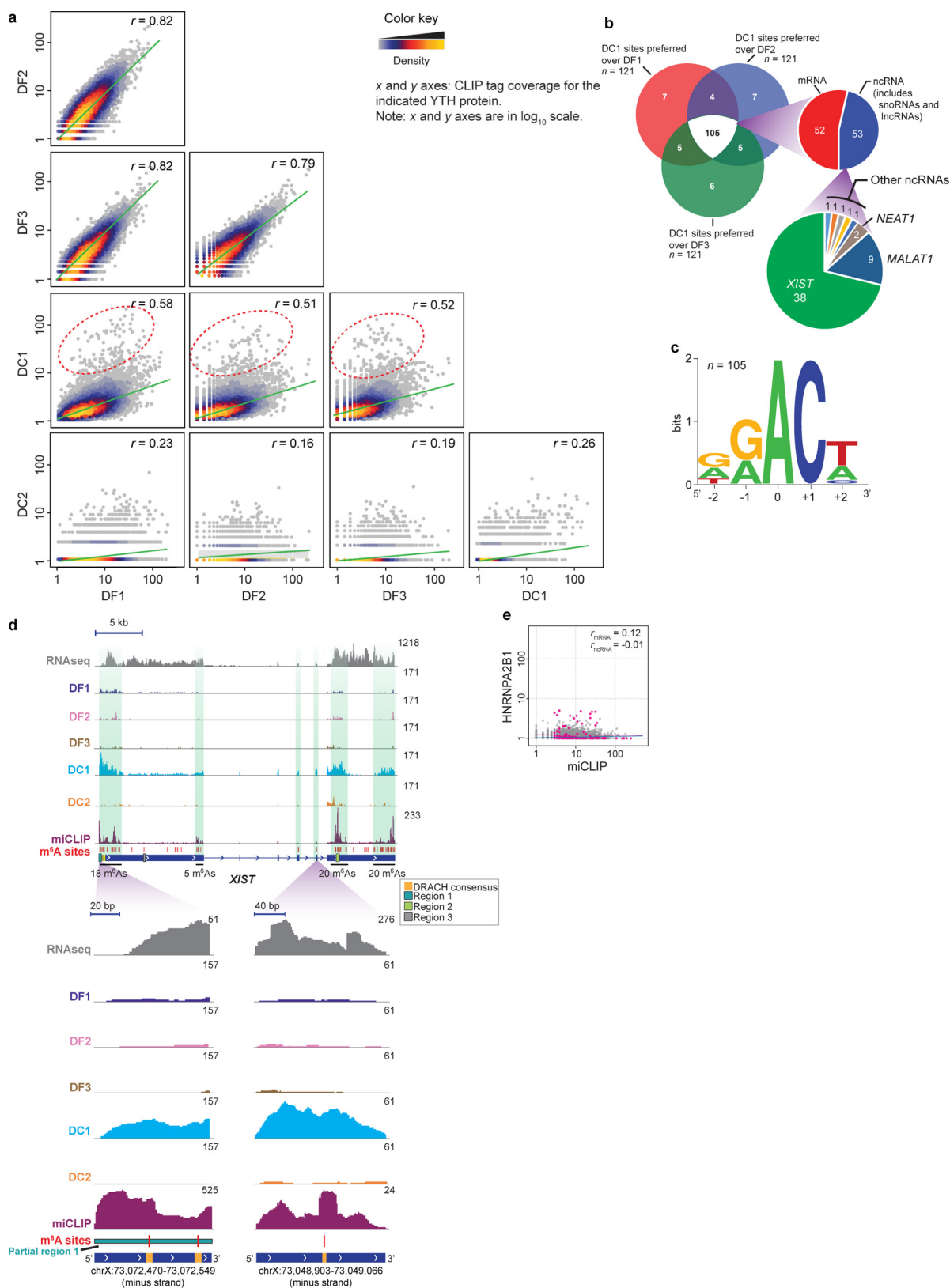
pulldown is shown. Antibodies and their antigenic peptide regions on the target proteins are provided in Supplementary Table 1. siRNA and shRNA target sequences in mRNA are listed in Supplementary Table 9. **h–l**, Autoradiograms from the nitrocellulose blots of samples used for each iCLIP library replicate. For each YTH protein, four biological replicates (rep1–4) were prepared. The red arrow confirms the position of the YTH protein after high RNase treatment and matches the size seen in **c–g**. Typically, UV crosslinking causes an increase in the intensity of the ^{32}P signal at the expected size of the YTH proteins (red arrow), indicating the formation of RNA–protein conjugates (lane 1 versus 3 in all panels). In the case of DC1, there is some ^{32}P signal even in the absence of UV crosslinking (lane 1 versus 3 in **k**). This type of background signal is due to autophosphorylation activity of the protein or of a co-immunoprecipitating protein kinase that phosphorylates DC1. RNase-sensitive smears were obtained for all of the YTH proteins (compare lanes 4–7 to lane 3 in **h–l**). Experiments using protein A/G beads that did not include the antibody (lane 2) did not show any signal in the region of interest. Overall, all the replicates of each YTH protein show highly specific RNA–protein conjugates of expected size with a minimal contamination of RNA–protein conjugates of other sizes. The eluted RNA material was used for constructing iCLIP libraries. Shown below the autoradiograms are western-blot loading controls (GAPDH and each YTH protein) as well as the controls that confirm the presence of the YTH protein in the immunoprecipitates.



Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | Comparison of transcriptome-wide RNA-binding sites of endogenous YTH proteins by iCLIP. **a–e**, YTH iCLIP library replicate reproducibility. For each YTH protein (DF1, DF2, DF3, DC1 and DC2), four independent biological replicate iCLIP libraries were constructed. Reproducibility of iCLIP tag coverage was assessed as in Extended Data Fig. 1f, g. Normalized iCLIP tag counts in uTPM from different replicates were compared on a scatter plot and the Pearson correlation coefficient (r) was determined. Scatter plots comparing mean tag counts of rep1 and rep2 (x axis), and rep3 and rep4 (y axis) are shown (left). A similar analysis was carried out for pairwise comparison of all the iCLIP replicates. The obtained correlation coefficients are shown on the heat map (right). The colour of the tiles in the heat map indicates the r value. YTH iCLIP replicates show similar and highly reproducible iCLIP tag coverage. The diagonal dashed line represents reference trend line for a perfect correlation ($r=1$, $x=y$). **f**, Enriched motifs for each YTH protein based on transcriptome-wide iCLIP binding data. Motif analysis of the binding sites recognized by DF1, DF2, DF3 and DC1 proteins in the iCLIP data showed a DRACH sequence as the most prominent motif, which matches the known consensus motif for m⁶A in the transcriptome⁴⁷. DC2 also showed the DRACH motif, as well as other motifs, which probably reflects its numerous RNA-binding domains (Extended Data Fig. 6a). DC1 predominantly bound DRACH in various transcriptomic and genomic features (Supplementary Table 10). These data suggest YTH proteins bind m⁶A in cells. RNA-binding motifs were identified using MEME analysis (see Methods). The percentage of analysed sites containing the identified motifs is shown in the top right. P values were obtained using the MEME CentriMo tool by a one-tailed binomial test. **g**, Global comparison of the distribution of YTH-binding sites and m⁶A sites on mRNA. We compared the metagene for each YTH protein binding site to the previously reported metagene of single-nucleotide

resolution miCLIP-identified m⁶A sites on mRNA¹⁷ (YTH protein, green; miCLIP, orange). Each curve represents a kernel density (y axis) plot of CITS distribution on a virtual transcript (x axis). Transcription start site, 5' UTR, start codon (AUG), CDS, stop codon, and 3' UTR are indicated on the virtual transcript. Vertical dashed lines mark UTR-CDS boundaries. Owing to a small number of DC2 sites that map to mRNA, a metagene for DC2 is not shown. **h**, Pairwise comparison of YTH iCLIP tag coverage at m⁶A sites shows distinct binding preferences for DC1. To determine whether YTH proteins recognize similar m⁶A sites on the basis of iCLIP tag coverage, we estimated the correlation coefficient of iCLIP tag coverage at each of 11,530 m⁶A sites in the transcriptome from a pairwise comparison of two YTH iCLIP libraries at miCLIP-identified m⁶A sites. The Pearson correlation coefficients are shown as a heat map. To identify YTH proteins that show similar binding preferences, libraries were hierarchically clustered based on the obtained correlation coefficients (see dendrogram below the heat map). This indicates that the DF proteins cluster together and show a similar binding pattern, and these proteins target similar m⁶A sites. Both DC1 and DC2 have atypical m⁶A site preferences; DC2 has a weak correlation with known m⁶A sites. Scatter plots used for this comparison are shown in Extended Data Fig. 8a. **i**, Genomic distribution of RNA-binding sites. To determine the genomic distribution of preferred YTH protein-RNA-binding sites, the statistically significant top 1,000 iCLIP CITS ($P < 0.0001$) with the highest iCLIP tag coverage (in uTPM) were mapped to different features of the human genome. DC1 exhibits prominent binding to ncRNAs (19% of top thousand CITS), while less than 2% of the DF1, DF2 or DF3 CITS lie within annotated ncRNAs, including lncRNAs. Most DF1-, DF2- and DF3-binding sites are located in mRNAs and introns. DC2 had negligible coverage in mRNAs, and predominantly bound to introns and intergenic sequences.

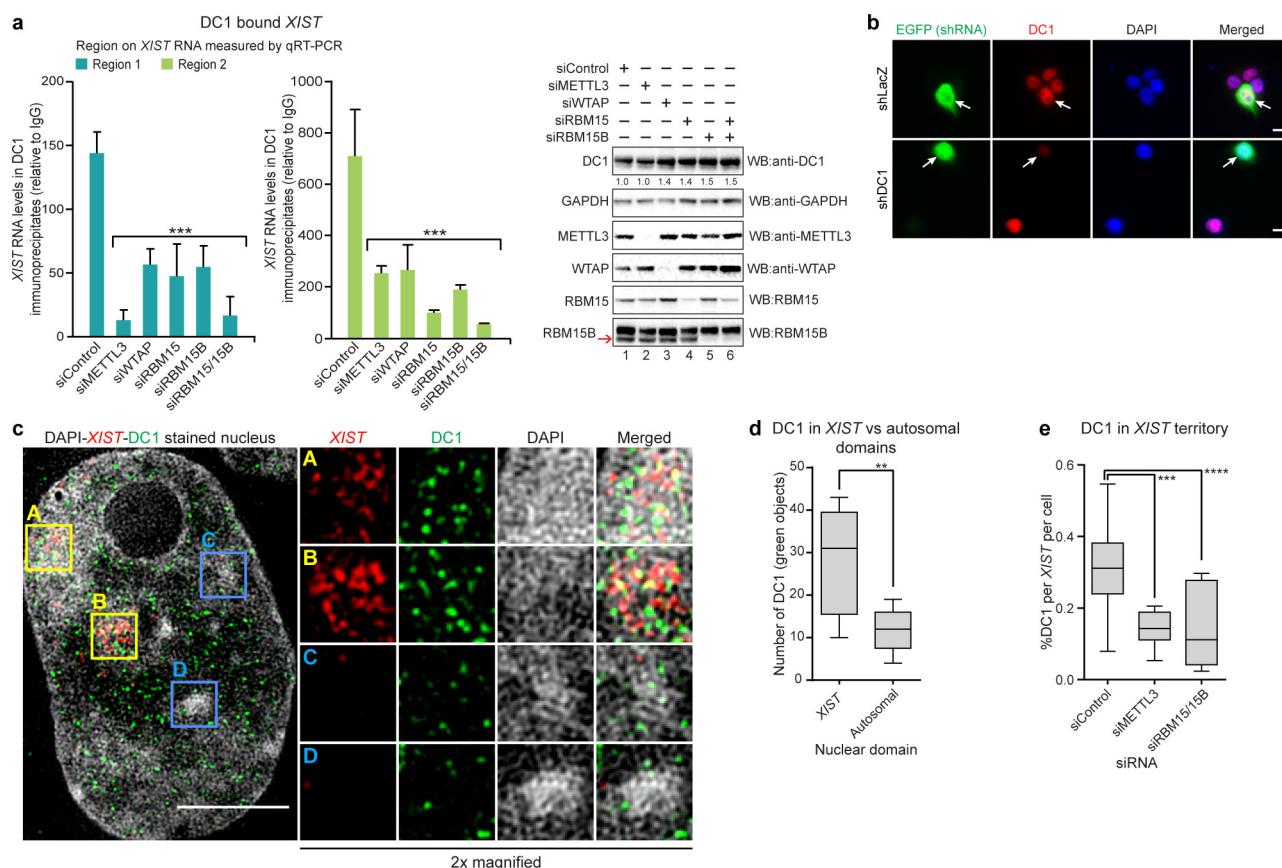


Extended Data Figure 8 | See next page for caption.

Extended Data Figure 8 | DC1 preferentially binds to a subset of m⁶A sites that are primarily localized to *XIST* and other ncRNAs.

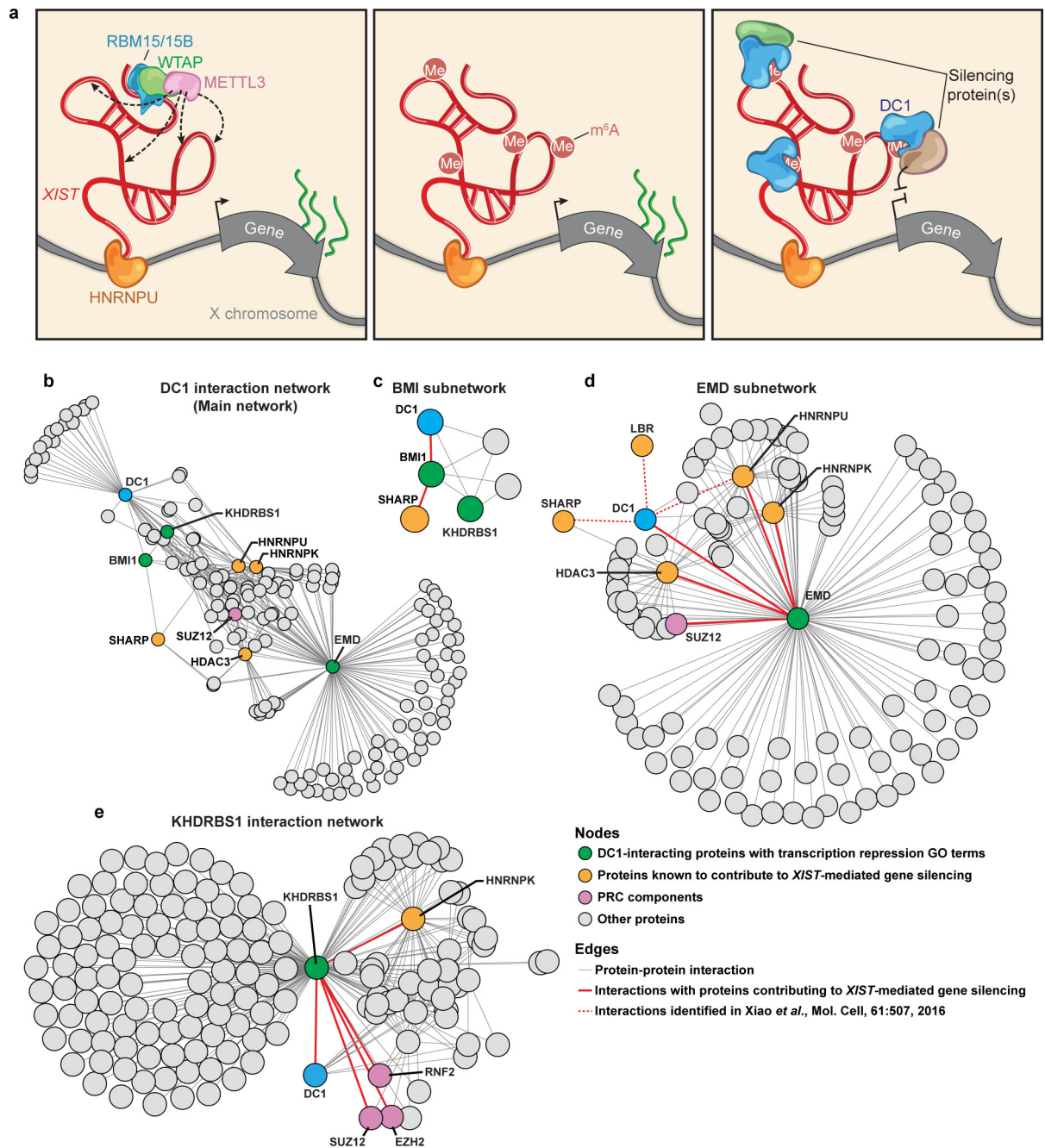
a, Pairwise comparison of YTH iCLIP libraries, and identification of DC1 preferred m⁶A sites. Shown are data used to generate the heat map in Extended Data Fig. 7h. In each pairwise analysis, two YTH proteins were compared for their binding to each m⁶A residue using normalized tag counts (see Methods), providing an estimate of the preferred binding partner for each m⁶A site for each YTH protein comparison. Tag counts in a window surrounding each m⁶A genomic coordinate (10 bp upstream and downstream) were determined for each YTH protein. Scatter plots are shown for each pair of indicated YTH proteins. m⁶A sites are plotted as points in which *x* and *y* coordinates represent the tag counts in the compared libraries. The DF family of proteins show highly similar binding preferences as indicated by their high Pearson correlation coefficients (*r*, top right corner of each plot). Hierarchical clustering as shown in Extended Data Fig. 7h supports the overall relatedness of the binding preferences of DF proteins. However, DC1 and DC2 show a pattern different from the DF proteins. DC2 shows low tag coverage on most m⁶A sites, and thus yields low *r* values. Notably, DC1 shows a global de-enrichment of binding at DF1, DF2 and DF3-preferred sites as seen by the flattened trend line (green). Additionally, DC1 shows enrichment at a unique set of m⁶A sites (the 1% of sites furthest from the trend line is highlighted with a red dashed ellipse in the comparison between DC1 and DF1, DF2 and DF3). **b**, A Venn diagram showing the number of sites preferred for DC1 over DF1, DF2 and DF3. The vast majority (105, white

shaded area) are the same between each comparison, meaning these sites are preferred by DC1 over any DF protein. The rightward projection shows that most of these m⁶A sites are in ncRNA, constituted primarily of *XIST* m⁶A sites. **c**, Sequence logo analysis shows that the DC1-preferred m⁶A sites conform to the DRACH-like m⁶A consensus motif seen throughout the transcriptome, not in a novel DC1-specific motif. **d**, Zoomed-in views of iCLIP tag distribution on *XIST* for the five YTH proteins on *XIST*. The miCLIP tag distribution also identifies regions enriched in m⁶A. Only DC1 (blue) exhibits prominent iCLIP tags on *XIST*, the other YTH proteins do not. Vertical green shading marks the regions of *XIST* that contain the highest density of m⁶A sites. RNA-seq reads are shown in read counts, iCLIP and miCLIP tags are shown in uTPM. Regions 1 and 2 contain RBM15/15B-binding sites, region 3 does not. These sites are indicated by coloured boxes. DC1 shows a higher number of iCLIP tags at regions 1 and 2, areas containing several m⁶A sites. Although region 3 (grey) shows a putative m⁶A site, DC1 shows poor binding, possibly owing to the structural organization of *XIST*. **e**, HNRNPA2B1 does not bind m⁶A sites on *XIST*. HNRNPA2B1 was previously shown to bind m⁶A sites on primary micro RNA (pri-miRNA) transcripts³⁵. We compared HNRNPA2B1 HITS-CLIP and miCLIP¹⁷ tag coverage (± 10 bp in uTPM) at 11,530 annotated m⁶A sites, and determined correlation coefficients for m⁶A sites in mRNA (red) and in ncRNA (blue). HNRNPA2B1 does not show any significant binding to m⁶A sites on mRNA and ncRNA. Notably, the miCLIP-identified m⁶A sites¹⁷ used in this analysis lacks m⁶A sites from pri-miRNAs.



Extended Data Figure 9 | DC1 binds *XIST* m⁶A in an METTL3-, RBM15-, and RBM15B-dependent manner. **a**, DC1 interacts with *XIST* in an RBM15/15B-dependent manner. Quantification of *XIST* in DC1 immunoprecipitates at regions 1 and 2 (left) by RNA immunoprecipitation followed by qPCR. Western blot analysis of protein from the siRNA-transfected cells (right). Knockdown of *METTL3*, *WTAP*, *RBM15* and *RBM15B* leads to a significant decrease in *XIST* enrichment from DC1 immunoprecipitates with *RBM15*/*RBM15B* double knockdown exhibiting the greatest decrease. These data indicate that DC1 binds *XIST* RNA in a METTL3/*RBM15*/15B-dependent manner. Region 3 showed no reproducible and detectable amplification, possibly owing to the poor binding of DC1. In Extended Data Fig. 8d, region 3 shows a very low DC1 iCLIP tag coverage. Data are mean \pm s.e.m. for three independent experiments. *** $P < 0.0001$ relative to *XIST* levels in siControl-transfected cells by unpaired two-sample *t*-test. **b**, Validation of DC1 antibody for immunofluorescence. Images of shLacZ- and shDC1-transfected HEK293T cells probed with DC1 antibody. DC1 exhibits a nuclear localization (red). In eGFP-expressing shDC1-transfected cells (arrow), the DC1 antibody signal is substantially lower than in a non-transfected cell in the same field (compare red signal, bottom row). Control knockdown with shLacZ-expressing plasmid shows DC1 staining similar to the non-transfected cells in the same view (red channel, arrow). Nuclei were stained with DAPI. Scale bars, 10 μ m. **c**, **d**, DC1 preferentially localizes to the *XIST* subnuclear compartment. 3D-SIM was used to examine the levels of DC1 in the *XIST* subnuclear compartment

compared to an autosomal domain in HEK293T cells following DC1 immunofluorescence labelling and *XIST* RNA-FISH. HEK293T cells are triploid, and thus exhibit two inactive X chromosomes^{48,49}. Left, a representative image showing DC1 (green), *XIST* (red) and DAPI (nucleus, grey-white) staining. Right, 2 \times magnification of highlighted regions (squares). DC1 is enriched in the *XIST* domains over similar dense autosomal compartments (right, top two versus bottom two rows). A distribution analysis of 3D-object counts performed on the DC1 signal in the *XIST* and autosomal domains also shows a significant enrichment (**d**, number of nuclei = 5, total *XIST* domains = 10, total autosomal domains = 10). Regions A and B (yellow squares) highlight two DAPI-stained inactivated X-chromosome territories marked by the presence of *XIST* (red). Areas C and D (blue squares) mark DAPI-stained autosomal domains. Scale bar, 5 μ m. In **d**, ** $P = 0.0023$ using two-tailed Mann-Whitney test. **e**, Localization of DC1 in *XIST* territory is METTL3- and RBM15/15B-dependent. To determine whether DC1 localizes to the *XIST* subnuclear compartment in an m⁶A-dependent manner, the number of DC1 spots in the *XIST* domain after METTL3 and RBM15/*RBM15B* knockdown was assessed by 3D-SIM, followed by image analysis. Knockdown of METTL3 and RBM15/*RBM15B* led to a significant decrease in the *XIST*-localized DC1. Box plot shows distribution of percentage of DC1 molecules (green objects) in *XIST* domain from the different knockdown cells. 10 nuclei per knockdown; *** $P = 0.0011$, **** $P = 0.0147$ relative to control knockdown in a two-tailed Mann-Whitney test.



Extended Data Figure 10 | See next page for caption.

Extended Data Figure 10 | Model for the role of m⁶A in *XIST*-mediated transcriptional silencing and DC1 protein–protein interaction network analysis.

a, A model for m⁶A-dependent *XIST*-mediated gene silencing. RBM15/RBM15B is the portion of the m⁶A methylation complex (that is, RBM15/RBM15B–WTAP–METTL3) that binds *XIST*. This binding enables methylation of adjacent adenosine residues in DRACH consensus sites. The m⁶A residues act as recruitment sites for DC1, which may facilitate and stabilize the assembly of silencing proteins on *XIST*. **b**, Protein–protein interaction (PPI) network analysis identifies a multi-component pathway that might mediate efficient *XIST*-mediated gene silencing. DC1 has no known protein domain that could directly mediate repression of gene transcription. We mined the PINA2 database²⁸ for the PPI network of DC1, as well as for proteins that interact with DC1-binding proteins and proteins that regulate *XIST*-mediated gene silencing (SHARP, HDAC3, HNRNPK, HNRNPU, NCOR2 (also known as SMRT), LBR, PRC1, and PRC2). A network of proteins that interact with DC1 is shown, as are the interactions of these proteins (subnetworks). Proteins that are linked to *XIST*-mediated silencing are indicated in pink (the PRC components) or orange. **c–e**, Subnetworks showing the presence of proteins involved in transcription repression. Gene Ontology terms were filtered from the main network in **b**. In **c** the DC1–BMI subnetwork is shown. This interaction is based on co-immunoprecipitation of DC1

with BMI1, a component of the PRC complex required to maintain gene repression⁵⁰. BMI1 may recruit SHARP, which directly binds *XIST* and mediates the recruitment of HDAC3 on the X chromosome. The EMD (emerin) subnetwork shown in **d**, is significantly enriched in proteins involved in transcription repression (false discovery rate < 0.05, $P < 0.05$) (Supplementary Table 7). DC1 interacts with EMD⁵¹, which is linked to proteins that are known to be necessary for *XIST*-mediated gene silencing (interactions indicated with bold red lines). A separate analysis of DC1 co-immunoprecipitated proteins identified by tandem immunoprecipitation followed by mass spectrometry analysis also shows the presence of SHARP and LBR proteins (interactions indicated with red dotted lines). Protein-binding partners of another DC1-interacting protein, KHDRBS1 (ref. 52) is shown in **e**. KHDRBS1 (also known as SAM68) is a well-known transcriptional repressor. Here KHDRBS1 is shown to interact with PRC component proteins SUZ12, EZH2, and RNF2. SUZ12 and EZH2 are components of the PRC2/EED–EZH2 complex that mediates histone methylation at K9 and K27 residues, leading to transcriptional repression. KHDRBS1 also interacts with *XIST*⁵. RNF2 is a component of PRC1 complex. RNF2 has E3 ubiquitin-protein ligase activity that mediates monoubiquitination of Lys119 of histone H2A (H2AK119Ub). Components of PRC1/2 and are found to be enriched on the inactivated X chromosome⁴.

A radio-pulsing white dwarf binary star

T. R. Marsh¹, B. T. Gänsicke¹, S. Hümmerich^{2,3}, F.-J. Hambsch^{2,3,4}, K. Bernhard^{2,3}, C. Lloyd⁵, E. Breed¹, E. R. Stanway¹, D. T. Steeghs¹, S. G. Parsons⁶, O. Toloza¹, M. R. Schreiber⁶, P. G. Jonker^{7,8}, J. van Roestel⁸, T. Kupfer⁹, A. F. Pala¹, V. S. Dhillon^{10,11,12}, L. K. Hardy¹⁰, S. P. Littlefair¹⁰, A. Aungwerojwit¹³, S. Arjyotha¹⁴‡, D. Koester¹⁵, J. J. Bochinski¹⁶, C. A. Haswell¹⁶, P. Frank² & P. J. Wheatley¹

White dwarfs are compact stars, similar in size to Earth but approximately 200,000 times more massive¹. Isolated white dwarfs emit most of their power from ultraviolet to near-infrared wavelengths, but when in close orbits with less dense stars, white dwarfs can strip material from their companions and the resulting mass transfer can generate atomic line² and X-ray³ emission, as well as near- and mid-infrared radiation if the white dwarf is magnetic⁴. However, even in binaries, white dwarfs are rarely detected at far-infrared or radio frequencies. Here we report the discovery of a white dwarf/cool star binary that emits from X-ray to radio wavelengths. The star, AR Scorpii (henceforth AR Sco), was classified in the early 1970s as a δ -Scuti star⁵, a common variety of periodic variable star. Our observations reveal instead a 3.56-hour period close binary, pulsing in brightness on a period of 1.97 minutes. The pulses are so intense that AR Sco's optical flux can increase by a factor of four within 30 seconds, and they are also detectable at radio frequencies. They reflect the spin of a magnetic white dwarf, which we find to be slowing down on a 10^7 -year timescale. The spin-down power is an order of magnitude larger than that seen in electromagnetic radiation, which, together with an absence of obvious signs of accretion, suggests that AR Sco is primarily spin-powered. Although the pulsations are driven by the white dwarf's spin, they mainly originate from the cool star. AR Sco's broadband spectrum is characteristic of synchrotron radiation, requiring relativistic electrons. These must either originate from near the white dwarf or be generated *in situ* at the M star through direct interaction with the white dwarf's magnetosphere.

The brightness of AR Sco varies on a 3.56-h period (Fig. 1a); it was this that caused the δ -Scuti classification⁵. The scatter visible in Fig. 1a prompted us to take optical photometry with the high-speed camera ULTRACAM⁶. These data and follow-up observations taken in the ultraviolet and near-infrared (Extended Data Table 1) all show strong double-humped pulsations on a fundamental period of 1.97 min (Figs 2 and 3); the scatter in Fig. 1a is the result of the pulsations. Most unusually of all, an hour-long observation at radio frequencies with the Australia Telescope Compact Array (ATCA) also shows these pulsations (Figs 2d, e and 3d). The pulse fraction, $(f_{\max} - f_{\min}) / (f_{\max} + f_{\min})$, where f_{\max} and f_{\min} are the maximum and minimum flux in a selected wavelength range, exceeds 95% in the far ultraviolet (Fig. 2), and is still 10% at 9 GHz in the radio frequency range. Only in X-rays did we not detect pulses (pulse fraction <30% at 99.7% confidence). AR Sco's optical magnitude (g') varies by a factor of 20 in flux, from 16.9 mag at its faintest to 13.6 mag at its peak.

We acquired optical spectra that show a cool M-type main-sequence star (Extended Data Fig. 1) with absorption lines that change in

radial velocity sinusoidally on the 3.56-h period with amplitude $K_2 = 295 \pm 4 \text{ km s}^{-1}$ (Fig. 1b; we use subscripts 1 and 2 to indicate the compact star and the M star, respectively). The 3.56-h period is therefore the orbital period of a close binary star. The M star's radial velocity amplitude sets a lower limit on the mass of its companion of $M_1 \geq (0.395 \pm 0.016)M_{\odot}$ (where M_{\odot} is the mass of the Sun). The compact object is not visible in the spectra, consistent with either a white dwarf or a neutron star, the only two types of object that can both support a misaligned magnetic dipole and spin fast enough to match the pulsations. The optical and ultraviolet spectra show atomic emission lines (Extended Data Figs 1 and 2) that originate from the side of the M star that faces the compact object (Extended Data Fig. 3). Their velocity amplitude relative to the M star sets a lower limit on the mass ratio $q = M_2/M_1 > 0.35$ (Extended Data Fig. 4). This, along with the requirement that the M star fits within its Roche lobe, defines mass ranges for each star of $0.81M_{\odot} < M_1 \lesssim 1.29M_{\odot}$ and $0.28M_{\odot} < M_2 \lesssim 0.45M_{\odot}$. The M star's spectral type (M5) suggests that its mass lies at the lower end of the allowed range for M_2 . Assuming that the M star is close to its Roche lobe, its brightness leads to a distance estimate $d = [M_2/(0.3M_{\odot})]^{1/3} (116 \pm 16) \text{ pc}$.

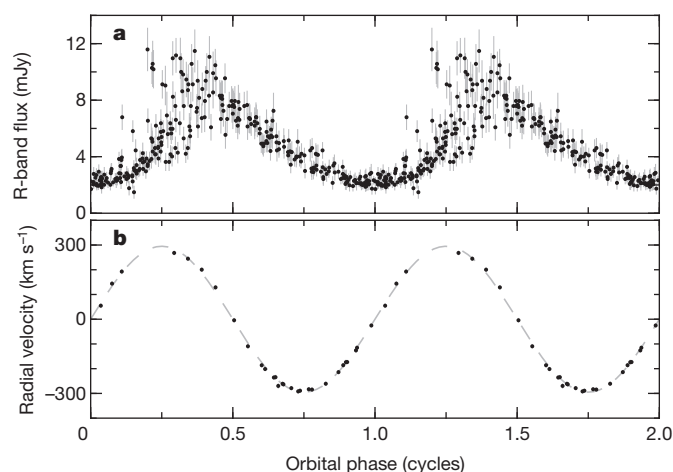


Figure 1 | AR Sco's optical brightness and radial velocity curve.

a, Photometry (30-s exposures) taken over 7 years shows a factor of four variation in brightness on a 3.56-h period, with large scatter at some phases. **b**, The M star varies sinusoidally in velocity on the same period, showing it to be the orbital period of a close, circular orbit binary star. The orbital phase is defined so that at phase 0 the M star is at its closest point to Earth. Error bars show $\pm 1\sigma$, but are too small to be seen in **b**.

¹Department of Physics, Gibbet Hill Road, University of Warwick, Coventry CV4 7AL, UK. ²Bundesdeutsche Arbeitsgemeinschaft für Veränderliche Sterne e.V. (BAV), Berlin, Germany. ³American Association of Variable Star Observers (AAVSO), Cambridge, Massachusetts, USA. ⁴Vereniging Voor Sterrenkunde (VVS), Brugge, Belgium. ⁵Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, UK. ⁶Instituto de Física y Astronomía, Universidad de Valparaíso, Avenida Gran Bretaña 1111, Valparaíso, Chile. ⁷SRON, Netherlands Institute for Space Research, Sorbonnelaan 2, 3584-CA Utrecht, The Netherlands. ⁸Department of Astrophysics/IMAPP, Radboud University Nijmegen, PO Box 9010, 6500 GL Nijmegen, The Netherlands. ⁹Division of Physics, Mathematics and Astronomy, California Institute of Technology, Pasadena, California 91125, USA. ¹⁰Department of Physics and Astronomy, University of Sheffield, Sheffield S3 7RH, UK. ¹¹Instituto de Astrofísica de Canarias (IAC), E-38205 La Laguna, Tenerife, Spain. ¹²Universidad de La Laguna, Departamento Astrofísica, E-38206 La Laguna, Tenerife, Spain. ¹³Department of Physics, Faculty of Science, Naresuan University, Phitsanulok 65000, Thailand. ¹⁴Program of Physics, Faculty of Science and Technology, Chiang Rai Rajabhat University, Chiang Rai 57100, Thailand. ¹⁵Institut für Theoretische Physik und Astrophysik, University of Kiel, 24098 Kiel, Germany. ¹⁶Department of Physical Sciences, The Open University, Milton Keynes, UK.

‡Deceased.

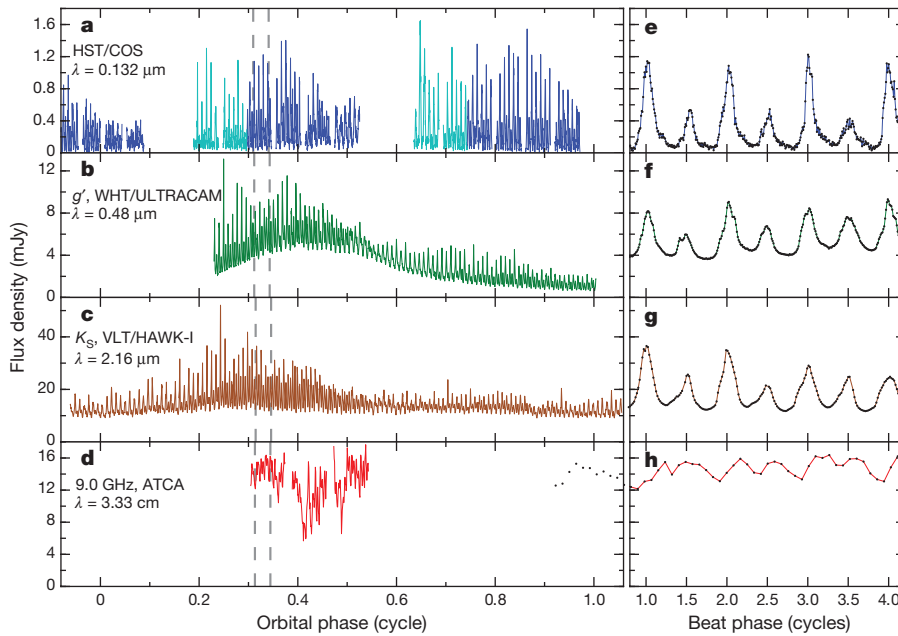


Figure 2 | Ultraviolet, optical, infrared and radio fluxes of AR Sco. **a–d**, High-speed measurements of the ultraviolet (**a**), optical (**b**), infrared (**c**) and radio fluxes (**d**) of AR Sco plotted against the orbital phase. **e–h**, An expanded view of sections of similar orbital phases (marked by dashed grey lines in **a–d**), is plotted against the beat pulsation phase. Black dots mark individual measurements. None of the four sets of data were taken simultaneously in time. The different colours in **a** indicate that the data were acquired in different orbital cycles.

The amplitude spectra of the pulsations show the presence of two components of similar frequency (Fig. 3). Using our own monitoring and archival optical data spanning seven years⁷, we measured precise values for the frequencies of these components, finding their difference to be within 20 parts per million of the orbital frequency, ν_O (Extended Data Figs 5 and 6, Extended Data Table 2). The natural interpretation is that the higher frequency component represents the spin frequency ν_S of the compact star (with a period $P_S = 1.95$ min), whereas its lower frequency and generally stronger counterpart is a reprocessed or ‘beat’ frequency $\nu_B = \nu_S - \nu_O$ ($P_B = 1.97$ min), assuming that the compact star spins in the same sense as the binary orbit.

AR Sco emits across the electromagnetic spectrum (Fig. 4, Extended Data Table 3) and, in the infrared and radio in particular, is orders of magnitudes brighter than the thermal emission from its component stars as represented by model atmospheres^{8,9} in Fig. 4. Integrating over the spectral energy distribution (SED) shown in Fig. 4 and adopting a distance of 116 pc, we find a maximum luminosity of about 6.3×10^{25} W and a mean of $\bar{L} \approx 1.7 \times 10^{25}$ W, well in excess of the combined luminosities of the stellar components $L_* \approx 4.4 \times 10^{24}$ W.

The two possible sources of this luminosity are the accretion power and spin-down power of the compact object. A spinning object of moment of inertia I loses energy at a rate $L_{\dot{\nu}} = -4\pi^2 I \nu_S \dot{\nu}_S$ where ν_S and $\dot{\nu}_S$ are the spin frequency and its time derivative. Using the archival optical data we found the spin frequency to be slowing, with a frequency derivative of $\dot{\nu}_S = (-2.86 \pm 0.36) \times 10^{-17}$ Hz s⁻¹. For masses M and radii R typical of neutron stars and white dwarfs ($M_{NS} = 1.4 M_\odot$, $R_{NS} = 10$ km; $M_{WD} = 0.8 M_\odot$, $R_{WD} = 0.01 R_\odot$), this leads to $L_{\dot{\nu}, NS} \approx 1.1 \times 10^{21}$ W and $L_{\dot{\nu}, WD} \approx 1.5 \times 10^{26}$ W. Compared with the mean luminosity excess (L_+) over the stellar contributions, $L_+ = \bar{L} - L_* \approx 1.3 \times 10^{25}$ W, this shows that the spin-down luminosity is sufficient to power the system if the compact object is a white dwarf, but not if it is a neutron star. Accretion is the only possible power source in the case of a neutron star—an accretion rate of $\dot{M}_{NS} \approx 1.0 \times 10^{-14} M_\odot$ yr⁻¹ suffices. Accretion could partially power a white dwarf, but it cannot be the main source because the rate required, $\dot{M}_{WD} \approx 1.3 \times 10^{-11} M_\odot$ yr⁻¹, is high enough that we should see Doppler-broadened emission lines from the accreting gas whereas AR Sco only shows features from the M star.

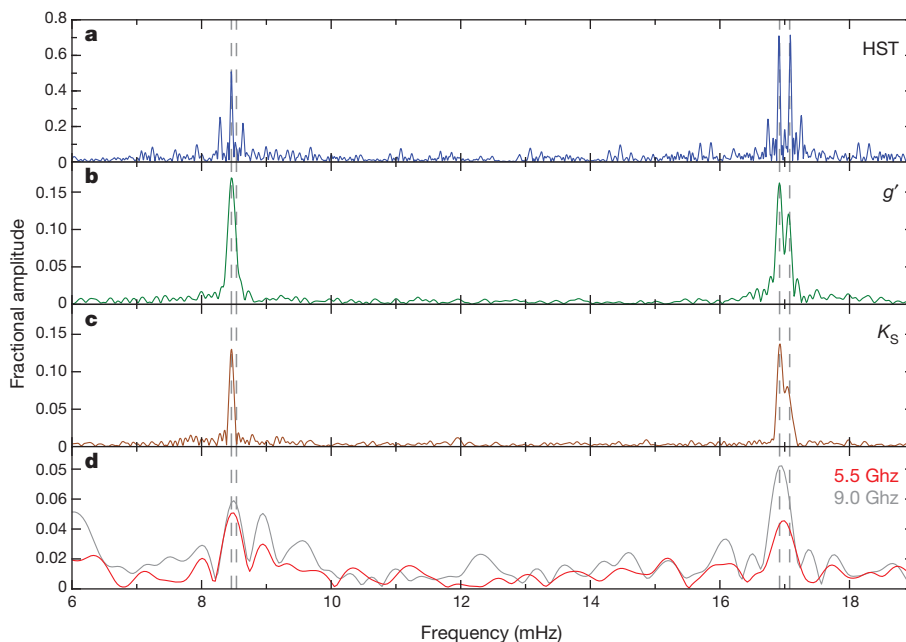


Figure 3 | Fourier amplitudes of the ultraviolet, optical, infrared and radio fluxes of AR Sco versus temporal frequency. **a–d**, Amplitude spectra corresponding to Fig. 2a–d. All bands show signals with a fundamental period of about 1.97 min (8.46 mHz) and its second harmonic. The signals have two components, clearest in the harmonic, which we identify as the spin frequency ν_S and beat frequency $\nu_B = \nu_S - \nu_O$, where ν_O is the orbital frequency. The pairs of grey dashed lines mark the positions of the beat (left) and spin (right) frequencies and their second harmonics. The beat component is the stronger of the two and defines the dominant 1.97-min pulsation period; the spin period is 1.95 min.

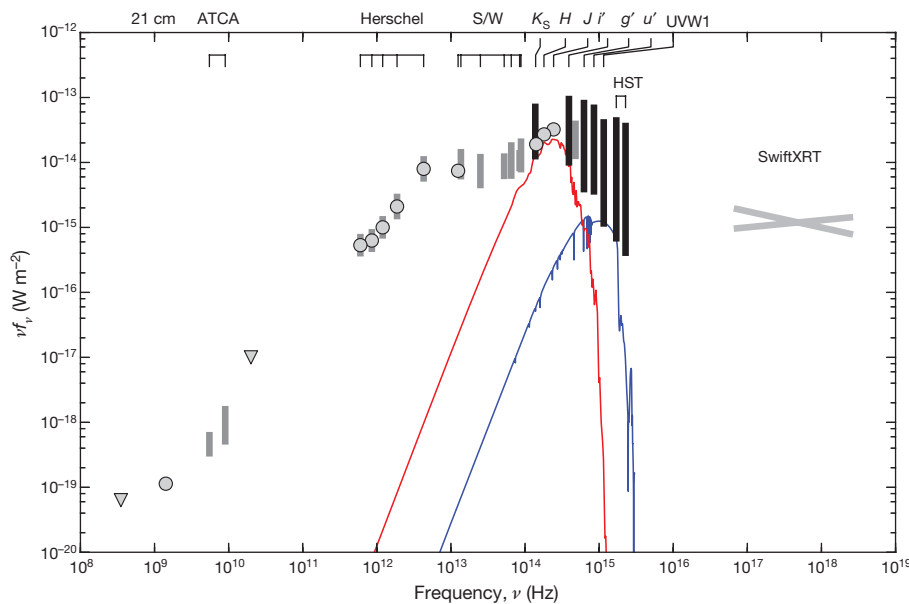


Figure 4 | The wide band SED of AR Sco. Black bars show the range spanned by intensive, time-resolved data; grey bars represent more limited datasets spanning less than the full variation. Grey points with error bars (1σ) represent single exposures. UVW1 refers to the Swift UVOT observations. The grey lines represent the $\pm 1\sigma$ range of the X-ray spectral slopes. The triangles are the upper-limits. S/W marks data from Spitzer and WISE. The solid lines show model atmospheres, extended at long wavelengths with blackbody spectra, for the M star (red, $R_2 = 0.36R_\odot$, $T_2 = 3,100$ K) and white dwarf (blue, $R_1 = 0.01R_\odot$, $T_2 = 9,750$ K upper limit) at a distance $d = 116$ pc. See Extended Data Tables 1 and 3 for details of the data sources.

The observations point towards a white dwarf as the compact object. First, AR Sco's distance of 116 pc is an order of magnitude closer than the nearest accreting neutron star known, Cen X-4¹⁰, but typical of white dwarf/main-sequence binaries (closer systems are known¹¹). Second, AR Sco's X-ray luminosity, $L_X \approx 4.9 \times 10^{23}$ W, is only 4% of the largely optical L_+ value. By contrast, the X-ray luminosities of accreting neutron stars are typically 100 times their optical luminosities¹². Third, at $P_S = 1.95$ min, AR Sco has a spin period that is an order of magnitude longer than any (neutron star-powered) radio pulsar known¹³. Finally, the upper limit masses $M_1 = 1.29M_\odot$ and $M_2 = 0.45M_\odot$ are simultaneously low for a neutron star but high for an M5 star. A $0.8M_\odot$ white dwarf with a $0.3M_\odot$ M dwarf is a more natural pairing.

The observed properties of AR Sco are unique. It may represent an evolutionary stage of a class of stars known as intermediate polars, which feature spinning magnetic white dwarfs accreting from low-mass stars in close binaries¹⁴. Only one intermediate polar, AE Aquarii (AE Aqr), has a broadband SED similar to AR Sco¹⁵ and comparably strong radio emission¹⁶, although it shows no radio pulsations¹⁷ ($<0.8\%$) and its 0.4% optical pulsations are much lower than the 70% in AR Sco. With a 25% pulse fraction, even the intermediate polar with the strongest-known optical pulsations, FO Aquarii¹⁸, falls well short of AR Sco. A key difference is perhaps the lack of substantial accretion in AR Sco compared with the intermediate polars. This can be seen from its X-ray luminosity, which is less than 1% of the X-ray luminosity of a typical intermediate polar¹⁹, but more obviously from its optical and ultraviolet emission lines, which come entirely from the irradiated face of the M star. Intermediate polars, by contrast, show Doppler-broadened line emission, often from accretion disks, and even AE Aqr (which is in an unusual 'propeller' state in which transferred matter is expelled on encountering the magnetosphere of its rapidly spinning $P_S = 33$ s white dwarf^{20,21}) shows broad and stochastically variable line emission. We can find no analogue of AR Sco's radio properties. Pulsed radio emission has been detected from brown dwarfs and M stars^{22,23}, but the broadband nature of AR Sco's emission, its short pulsation period and lack of circular polarization (our ATCA data constrain it to $<10\%$), distinguish it from these sources.

The white dwarf in AR Sco is currently spinning down on a timescale $\tau = \frac{\nu}{\dot{\nu}} = 10^7$ yr. White dwarfs are not born spinning rapidly²⁴, and a previous stage of accretion-driven spin-up is required. Depending on the distance at which the accreting material coupled to the white dwarf's magnetic field, between $0.002M_\odot$ and $0.015M_\odot$ of matter are required to reach $P_S = 1.95$ min. For an accretion rate of $10^{-9}M_\odot \text{ yr}^{-1}$, typical of systems with similar periods, this takes from 2×10^7 yr to

1.5×10^8 yr. Both spin-up and spin-down timescales are much shorter than the probable age of the system: the cooling age of the white dwarf alone exceeds 1.2×10^9 yr (ref. 25). We could therefore be seeing one of many such episodes in AR Sco's history. There is empirical evidence for similar cycling of the accretion rate in both white dwarf^{26,27} and neutron star binary systems^{28,29}. If so, because the spin-up and spin-down timescales are of similar magnitude, there would be a good chance of catching the spin-down phase.

AR Sco's extremely broadband SED is indicative of synchrotron emission from relativistic electrons. A large fraction seems to come from the cool M star. We infer this from the dominant beat frequency component that in the absence of accretion can only come from the M star. Because the M star occupies approximately 1/40 of the sky as seen from the white dwarf, whereas the spin-down luminosity is about 11.5 times the mean electromagnetic power, a mechanism is required for the transfer of energy from the white dwarf to the M dwarf that is more than $40/11.5 = 3.5$ times more efficient than the interception of isotropically emitted radiation. At the same time, direct pulsed emission from the white dwarf must not overwhelm the beat component. Two possibilities are collimated fast particle outflows and the direct interaction of the white dwarf's magnetosphere with the M dwarf, but the exact emission mechanism operating in AR Sco is perhaps its most mysterious feature.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 31 March; accepted 25 May 2016.

Published online 27 July 2016.

- Althaus, L. G., Córscico, A. H., Isern, J. & García-Berro, E. Evolutionary and pulsational properties of white dwarf stars. *Astron. Astrophys. Rev.* **18**, 471–566 (2010).
- Szkody, P. et al. Cataclysmic variables from the Sloan Digital Sky Survey. VIII. The final year (2007–2008). *Astron. J.* **142**, 181–189 (2011).
- Revnivtsev, M., Sazonov, S., Krivonos, R., Ritter, H. & Sunyaev, R. Properties of the galactic population of cataclysmic variables in hard X-rays. *Astron. Astrophys.* **489**, 1121–1127 (2008).
- Parsons, S. G. et al. A magnetic white dwarf in a detached eclipsing binary. *Mon. Not. R. Astron. Soc.* **436**, 241–252 (2013).
- Satyvaldiev, V. On seventeen variable stars. *Astron. Tsirk.* **633**, 7–8 (1971).
- Dhillon, V. S. et al. ULTRACAM: an ultrafast, triple-beam CCD camera for high-speed astrophysics. *Mon. Not. R. Astron. Soc.* **378**, 825–840 (2007).
- Drake, A. J. et al. First results from the Catalina Real-Time Transient Survey. *Astron. J.* **136**, 870–884 (2009).
- Koester, D. White dwarf spectra and atmosphere models. *Mem. Soc. Astron. Ital.* **81**, 921–931 (2010).
- Husser, T.-O. et al. A new extensive library of PHOENIX stellar atmospheres and synthetic spectra. *Astron. Astrophys.* **553**, A6 (2013).

10. Chevalier, C., Ilovaisky, S. A., van Paradijs, J., Pedersen, H. & van der Klis, M. Optical studies of transient low-mass X-ray binaries in quiescence. I — Centaurus X-4: orbital period, light curve, spectrum and models for the system. *Astron. Astrophys.* **210**, 114–126 (1989).
 11. Thorstensen, J. R., Lépine, S. & Shara, M. Parallax and distance estimates for twelve cataclysmic variable stars. *Astron. J.* **136**, 2107–2114 (2008).
 12. Bradt, H. V. D. & McClintock, J. E. The optical counterparts of compact galactic X-ray sources. *Annu. Rev. Astron. Astrophys.* **21**, 13–66 (1983).
 13. Manchester, R. N., Hobbs, G. B., Teoh, A. & Hobbs, M. The Australia Telescope National Facility Pulsar Catalogue. *Astron. J.* **129**, 1993–2006 (2005).
 14. Patterson, J. The DQ Herculis stars. *Publ. Astron. Soc. Pacif.* **106**, 209–238 (1994).
 15. Oruru, B. & Meintjes, P. J. X-ray characteristics and the spectral energy distribution of AE Aquarii. *Mon. Not. R. Astron. Soc.* **421**, 1557–1568 (2012).
 16. Bookbinder, J. A. & Lamb, D. Q. Discovery of radio emission from AE Aquarii. *Astrophys. J.* **323**, L131–L135 (1987).
 17. Bastian, T. S., Beasley, A. J. & Bookbinder, J. A. A search for radio pulsations from AE Aquarii. *Astrophys. J.* **461**, 1016–1020 (1996).
 18. Patterson, J. & Steiner, J. E. H2215–086 — King of the DQ Herculis stars. *Astrophys. J.* **264**, L61–L64 (1983).
 19. Pretorius, M. L. & Mukai, K. Constraints on the space density of intermediate polars from the Swift-BAT survey. *Mon. Not. R. Astron. Soc.* **442**, 2580–2585 (2014).
 20. Wynn, G. A., King, A. R. & Horne, K. A magnetic propeller in the cataclysmic variable AE Aquarii. *Mon. Not. R. Astron. Soc.* **286**, 436–446 (1997).
 21. Meintjes, P. J. & Venter, L. A. The diamagnetic blob propeller in AE Aquarii and non-thermal radio to mid-infrared emission. *Mon. Not. R. Astron. Soc.* **360**, 573–582 (2005).
 22. Berger, E. *et al.* Discovery of radio emission from the brown dwarf LP944–20. *Nature* **410**, 338–340 (2001).
 23. Hallinan, G. *et al.* Periodic bursts of coherent radio emission from an ultracool dwarf. *Astrophys. J.* **663**, L25–L28 (2007).
 24. Charpinet, S., Fontaine, G. & Brassard, P. Seismic evidence for the loss of stellar angular momentum before the white-dwarf stage. *Nature* **461**, 501–503 (2009).
 25. Renedo, I. *et al.* New cooling sequences for old white dwarfs. *Astrophys. J.* **717**, 183–195 (2010).
 26. Hessman, F. V., Gänsicke, B. T. & Mattei, J. A. The history and source of mass-transfer variations in AM Herculis. *Astron. Astrophys.* **361**, 952–958 (2000).
 27. Manser, C. J. & Gänsicke, B. T. Spectroscopy of the enigmatic short-period cataclysmic variable IR Com in an extended low state. *Mon. Not. R. Astron. Soc.* **442**, L23–L27 (2014).
 28. Archibald, A. M. *et al.* A radio pulsar/X-ray binary link. *Science* **324**, 1411–1414 (2009).
 29. Papitto, A. *et al.* Swings between rotation and accretion power in a binary millisecond pulsar. *Nature* **501**, 517–520 (2013).
- Acknowledgements** T.R.M., E.R.S., D.S., E.B., P.J.W., V.S.D., S.P.L. and ULTRACAM were supported by the Science and Technology Facilities Council (STFC, grant numbers ST/L000733 and ST/M001350/1). B.T.G., A.P. and P.G.J. acknowledge support from the European Research Council (ERC, grant numbers 320964 and 647208). O.T., S.G.P. and M.R.S. acknowledge support from Fondecyt (grant numbers 3140585 and 1141269). M.R.S. also received support from Millenium Nucleus RC130007 (Chilean Ministry of Economy). A.A. acknowledges support from the Thailand Research Fund (grant number MRG5680152) and the National Research Council of Thailand (grant number R2559B034). The analysis in this paper is based on observations collected with telescopes of the Isaac Newton Group in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias, the European Organisation for Astronomical Research in the Southern Hemisphere (observing programmes 095.D-0489, 095.D-0739, 095.D-0802), the NASA/ESA Hubble Space Telescope (observing programmes 14470) and the Thai National Telescope. Archival data from the Herschel, Spitzer and WISE space observatories, and from the Catalina Sky Survey were used. We thank the Swift mission PI for a target-of-opportunity program on AR Sco with the XRT and UVOT instruments and Jamie Stevens for carrying out the ATCA Director's Discretionary Time observations. This paper is dedicated to the memory of Sirinipa Arjyotha.
- Author Contributions** T.R.M. organised observations, analysed the data, interpreted the results and was the primary author of the manuscript. B.T.G., A.F.P., E.B., S.G.P., P.G.J., J.v.R., T.K., M.R.S. and O.T. acquired, reduced and analysed optical and ultraviolet spectroscopy. E.R.S. acquired, reduced and analysed the ATCA radio data. S.H., F.-J.H., K.B., C.L. and P.F. first identified the unusual nature of AR Sco and started the optical monitoring campaign. V.S.D., L.K.H., S.P.L., A.A., S.A., J.J.B. and C.A.H. acquired and reduced the high-speed optical photometry. D.T.S. and P.J.W. acquired and analysed Swift and archival X-ray data. D.K. calculated the white dwarf model atmosphere. All authors commented on the manuscript.
- Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.R.M. (t.r.marsh@warwick.ac.uk).
- Reviewer Information** *Nature* thanks S. Ransom and M. H. van Kerkwijk for their contribution to the peer review of this work.

METHODS

Data sources. AR Sco's location in the ecliptic plane, not far from the Galactic Centre and only 2.5° northwest of the centre of the Ophiuchus molecular cloud, means that it appears in many archival observations. It is detected in the FIRST 21 cm radio survey³⁰, the Two Micron All Sky Survey (2MASS)³¹, the Catalina Sky Survey (CSS)⁷ and in the Herschel, WISE and Spitzer infrared satellite archives^{32–34}. Useful upper limits come from non-detections in the Australia Telescope 20 GHz (AT20G) survey³⁵ and the WISH survey³⁶. Flux measurements, ranges (when time resolved data are available) and upper limits from these sources are listed in Extended Data Table 3.

We supplemented these data with our own intensive observations on a variety of telescopes and instruments: the 8.2 m Very Large Telescope (VLT) with the FORS and X-SHOOTER optical/NIR spectrographs and the HAWK-I NIR imager; the 4.2 m William Herschel Telescope (WHT) with the ISIS spectrograph and the ULTRACAM high-speed camera⁶; the 2.5 m Isaac Newton Telescope (INT) with the Intermediate Dispersion Spectrograph (IDS); the 2.4 m Thai National Telescope with the ULTRASPEC high-speed camera³⁷; the ultraviolet/optical and X-ray instruments UVOT and XRT on the Swift satellite; the COS UV spectrograph on the Hubble Space Telescope (HST); radio observations on the Australia Telescope Compact Array (ATCA). Optical monitoring data came from a number of small telescopes. We include here data taken with a 406 mm telescope at the Remote Observatory Atacama Desert (ROAD) in San Pedro de Atacama, Chile³⁸. Extended Data Table 1 summarizes these observations.

The orbital, spin and beat frequencies. The orbital, spin and beat frequencies were best measured from the small-telescope data because of the long time interval covered by these observations. For example, see the amplitude spectrum around the spin/beat components of the clear filter data from 19–28 July 2015 shown in Extended Data Fig. 5. The final frequencies, which give the dashed lines of Extended Data Fig. 5, were obtained from the CSS data. These consisted of 305 exposures each of 30 s duration spanning the interval 30 May 2006 to 8 July 2013. We rejected 6 points that lay more than 4σ from the multisinusoid fits that we now describe. To search for signals in these sparsely sampled data, we first transformed the UTC times of the CSS data to a uniform timescale (TDB) and then corrected these for light-travel delays to the Solar System barycentre. The periodogram of these data are dominated by the strong orbital modulation, which leaks so much power across the spectrum owing to the sparse sampling that the spin/beat component can only be seen after the orbital signal is removed. Once this was done, beat and spin components matching those of Extended Data Fig. 5 could be identified (Extended Data Fig. 6). We carried out bootstrap multisinusoid fits to compute the distributions of the orbital, beat and spin frequencies. The orbital frequency closely follows a Gaussian distribution; the beat and spin distributions are non-Gaussian in their high- and low-frequency wings, respectively, but are nevertheless well-defined. Statistics computed from these distributions are listed in Extended Data Table 2.

Having established that the two pulsation frequencies are separated by the orbital frequency, we carried out a final set of fits in which we enforced the relation $\nu_S - \nu_B = \nu_O$, but also allowed for a linear drift of the pulsation frequency to be sensitive to any change in the pulsation frequency. This led to a significant improvement in χ^2 ($>99.99\%$ significance on an F test) that dropped from 326 to 289 for the 299 fitted points relative to a model in which the frequencies did not vary (after scaling uncertainties to yield $\chi^2/N \approx 1$ for the final fit). Bootstrap fits gave a near-Gaussian distribution for the frequency derivative with $\dot{\nu} = (-2.86 \pm 0.36) \times 10^{-17} \text{ Hz s}^{-1}$.

Pulsations are detected at all wavelengths with suitable data other than X-rays, where limited signal (approximately 630 source photons in 10.2 ks) leads to the upper limit of a 30% pulse fraction quoted in the main text. The Swift X-ray observations were taken in 1,000 s chunks over the course of more than one month and we searched for the pulsations by folding into 20 bins and fitting a sinusoid to the result. There were no major signals on the beat or spin periods or their harmonics. We used a power-law fit to the X-ray spectrum to deduce the slopes shown in Fig. 4.

The M star's spectral type and distance. The CSS data establish the orbital period $P = 0.14853528(8)$ d, but not the absolute phase of the binary. This we derived from observations of the M star, which also led to a useful constraint on the distance to the system. The VLT+FORs data were taken shortly before the photometric minimum, allowing a clear view of the M star's contribution. We used M star spectral-type templates developed from SDSS spectra³⁹ to fit AR Sco's spectrum, applying a flux scaling factor α to the selected template and adding a smooth continuum to represent any extra flux in addition to the M star. The smooth spectrum was parameterized by $\exp(a_1 + a_2\lambda)$ to ensure positivity. The coefficients a_1 , a_2 and α were optimized for each template, with emission lines masked as they are not modelled by the smooth spectrum. Out of the templates available (M0–9 in unit steps), the M5 spectrum gave by far the best match with $\chi^2 = 24,029$

for 1,165 points fitted compared with $>100,000$ for the M4 and M6 templates on either side (Extended Data Fig. 1). The templates used were normalized such that the scaling factor $\alpha = (R_2/d)^2$. We found $\alpha = 3.02 \times 10^{-21}$, so $R_2/d = 5.5 \times 10^{-11}$. Assuming that the M star is close to its Roche lobe (there is evidence supporting this assumption in the form of ellipsoidal modulations of the minima between pulsations in the HAWK-I data, Fig. 2), its mean density is fixed by the orbital period, which means that its radius is fixed by its mass. Assuming $M_2 = 0.3 M_\odot$, for reasons outlined in the main text, we find that $R_2 = 0.36 R_\odot$, and hence $d = 149$ pc. This is an overestimate as the FORS spectrum was taken through a narrow slit. We estimated a correction factor by calculating the i' band flux of the spectrum (2.50 mJy) and comparing it to the mean i' band flux (4.11 mJy) of the ULTRACAM photometry over the same range of orbital phase. This is approximate given that the ULTRACAM data were not taken simultaneously with the FORS data and there may be stochastic variations in brightness from orbit to orbit, however, the implied 61% throughput is plausible given the slit width of $0.7''$ and atmospheric seeing of around $1''$. The final result is the distance quoted in the main text of $d = 116 \pm 16$ pc, and allows for uncertainties in the calibration of the surface brightness of the templates and in the slit-loss correction.

We used the radius, spectral type and distance to estimate the K_S flux density from the donor as $f_{K_S} = 9.4$ mJy. The minimum observed flux density from the HAWK-I data is 9.1 mJy. Uncertainties in the extrapolation required to estimate the K_S flux and from ellipsoidal modulations allow the numbers to be compatible, but they suggest that the estimated distance is as low as it can be and that the M star dominates the K_S flux at minimum light. The estimated M star fluxes for i' and g' , $f_{i'} = 1.79$ mJy and $f_{g'} = 0.07$ mJy, respectively, are well below the minimum observed fluxes of 2.57 mJy and 0.624 mJy in the same bands. We do not detect the white dwarf. The strongest constraint comes from the HST far ultraviolet data that at its lowest require $T_1 < 9,750$ K. A white dwarf model atmosphere of $T = 9,750$ K, $\log[g] = 8$, is plotted in Fig. 1 (corrected for slit-losses) and also (without slit losses) in Fig. 2, which shows the average HST spectrum. Given the small maximum contribution of the white dwarf seen in these figures, the absence of absorption features from the white dwarf is unsurprising.

The M star's radial velocity. We used spectra taken with the ISIS spectrograph on the WHT and X-SHOOTER on the VLT to measure radial velocities of the M star using the Na I 8,200 doublet lines. These vary sinusoidally on the same 3.56 h period as the slowest photometric variation (Fig. 1), hence our identification of this period as the orbital period. We fitted the velocities with

$$V_R = \gamma + K_2 \sin(2\pi(t - T_0)/P)$$

where V_R is the radial velocity. The period P was fixed at the value obtained from the CSS data, $P = 0.14853528$ d, and the systemic offset γ allowed to float free for each distinct subset of the data to allow for variable offsets. We found $K_2 = 295 \pm 4 \text{ km s}^{-1}$ and $T_0 = 57264.09615(33)$ d, thus the orbital ephemeris of AR Sco is

$$\text{BMJD(TDB)} = 57264.09615(33) + 0.14853528(8)E,$$

where BMJD is the Barycentric Modified Julian Day, E is the cycle number and the time scale is TDB, corrected to the barycentre of the Solar System, expressed as a Modified Julian Day number ($\text{MJD} = \text{JD} - 2400000.5$). This ephemeris is important in establishing the origin of the emission lines, as will be shown below.

The radial velocity amplitude and orbital period along with Kepler's third law define the 'mass function'

$$\frac{M_1^3 \sin^3 i}{(M_1 + M_2)^2} = \frac{PK_2^3}{2\pi G} = (0.395 \pm 0.016) M_\odot$$

where i is the orbital inclination. This is a hard lower limit to the mass of the compact object, M_1 , which is only met for $i = 90^\circ$ and $M_2 = 0$. There is, however, a caveat to this statement: it is sometimes observed that irradiation can weaken the absorption lines on the side of the cool star facing the compact object causing the observed radial velocity amplitude to be an overestimate of the true amplitude^{40,41}. If this effect applied here, which we suspect it might, both K and the mass function limit would need to be reduced. Given the large intrinsic variability of AR Sco, and the lack of flux-calibrated spectra, it was not possible to measure the absolute strength of Na I. We attempted therefore to search for the influence of irradiation from another side effect, which is that it causes the radial velocity to vary non-sinusoidally⁴². We failed to detect any obvious influence of irradiation through this method, but its effectiveness may be limited by the heterogeneous nature of our data, which required multiple systematic velocity offsets. Despite our failure to detect clear signs of the effect of irradiation on the M star's radial velocities, we would not be surprised if the true value of K was anything up to about 20 km s^{-1} lower than we measure. However, with no clear evidence for the effect, in this Letter we proceed on the basis that we have measured the true value of the radial velocity

amplitude at the M star's centre of mass. This is conservative in the sense that any reduction in K would move the mass limits we deduce to lower values, which would tilt the balance even more heavily towards a white dwarf as the compact star. The ultraviolet and optical emission lines come from the irradiated face of the M star and their amplitude compared with K_2 sets a lower limit to the relative size of the M star, and hence, through Roche geometry, the mass ratio $q = M_2/M_1$. Extended Data Fig. 4 shows how the emission measurements lead to the quoted limit of $q > 0.35$, which leads in turn to the lower limits $M_1 > 0.81M_\odot$ and $M_2 > 0.28M_\odot$ quoted in the main text.

The orbital period of a binary star sets a lower limit on the mean densities of its component stars⁴³. As the mean densities of main-sequence stars decrease with increasing mass, this implies that we can set an upper limit to the mass of any main-sequence component. In the case of AR Sco we find that $\langle \rho \rangle > 8,900 \text{ kg m}^{-3}$, which leads to $M_2 < 0.42M_\odot$; the slightly larger value of $0.45M_\odot$ quoted in the text allows for uncertainty in the models. The limit becomes an equality when the M star fills its Roche lobe, which we believe to be the case, or very nearly so, for AR Sco. However, we expect that even in this case the number deduced still functions as an upper limit because the mass-losing stars in close binaries are generally oversized and therefore less dense than main-sequence stars of the same mass⁴⁴. Indeed, systems with similar orbital periods to that of AR Sco have donor star masses in the range $0.2M_\odot$ to $0.3M_\odot$ (ref. 44). This, and the M5 spectral type, are why we favour a mass of $M_2 \approx 0.3M_\odot$, close to the lower limit on M_2 .

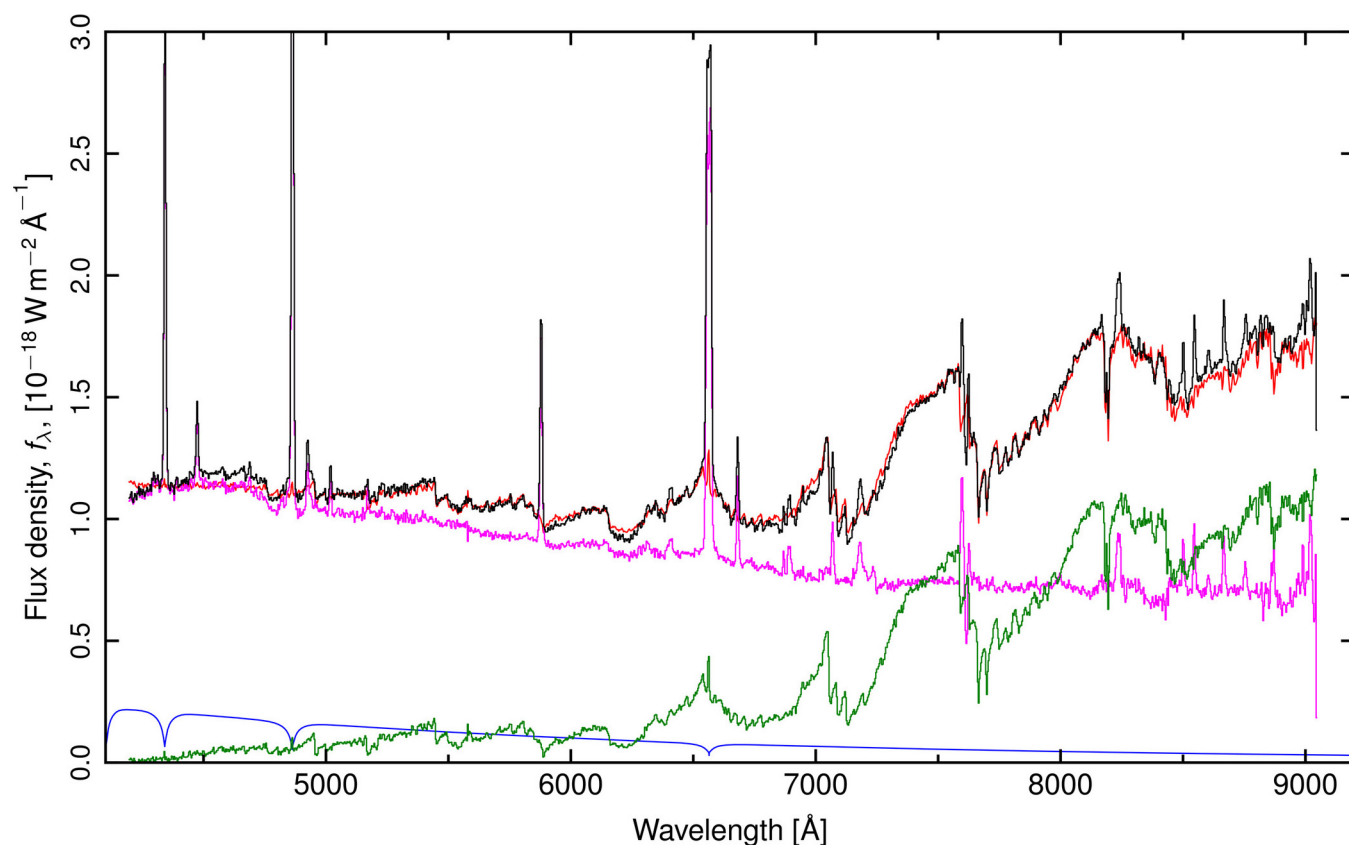
Brightness temperature at radio wavelengths. The pulsations in radio flux are a remarkable feature of AR Sco, unique amongst known white dwarfs and white dwarf binaries. If we assume that, as at other wavelengths and as suggested by the alignment of the second harmonic power with $2\nu_B$ (Extended Data Fig. 3), they arise largely from the M star, then we can deduce brightness temperatures from the relation

$$T_b = \frac{\lambda^2}{2\pi k} \left(\frac{d}{R_2} \right)^2 f_\nu$$

These work out to be approximately 10^{12} K and approximately 10^{13} K for the observations at $\nu = 5.5 \text{ GHz}$ and $\nu = 1.4 \text{ GHz}$ respectively. Although the value at the lowest frequency exceeds the limit (of approximately 10^{12} K) at which severe cooling of the electrons due to inverse Compton scattering is thought to occur⁴⁵, this is not necessarily a serious issue given the short-term variability exhibited by the source. The limits can be lowered by appealing to a larger emission region as the radio data in hand are not enough to be certain that emission arises solely on the M star. Even so, the 0.98 min second-harmonic pulsations that are seen in the radio flux suggest an upper limit to the size of the emission region of $25R_\odot$ from light-travel time alone. This implies a minimum brightness temperature of 10^9 K at 1.4 GHz , showing clearly that the radio emission is non-thermal in origin. We assume that synchrotron emission dominates; whereas there may be thermal and cyclotron components at shorter wavelengths, there is no clear evidence for either.

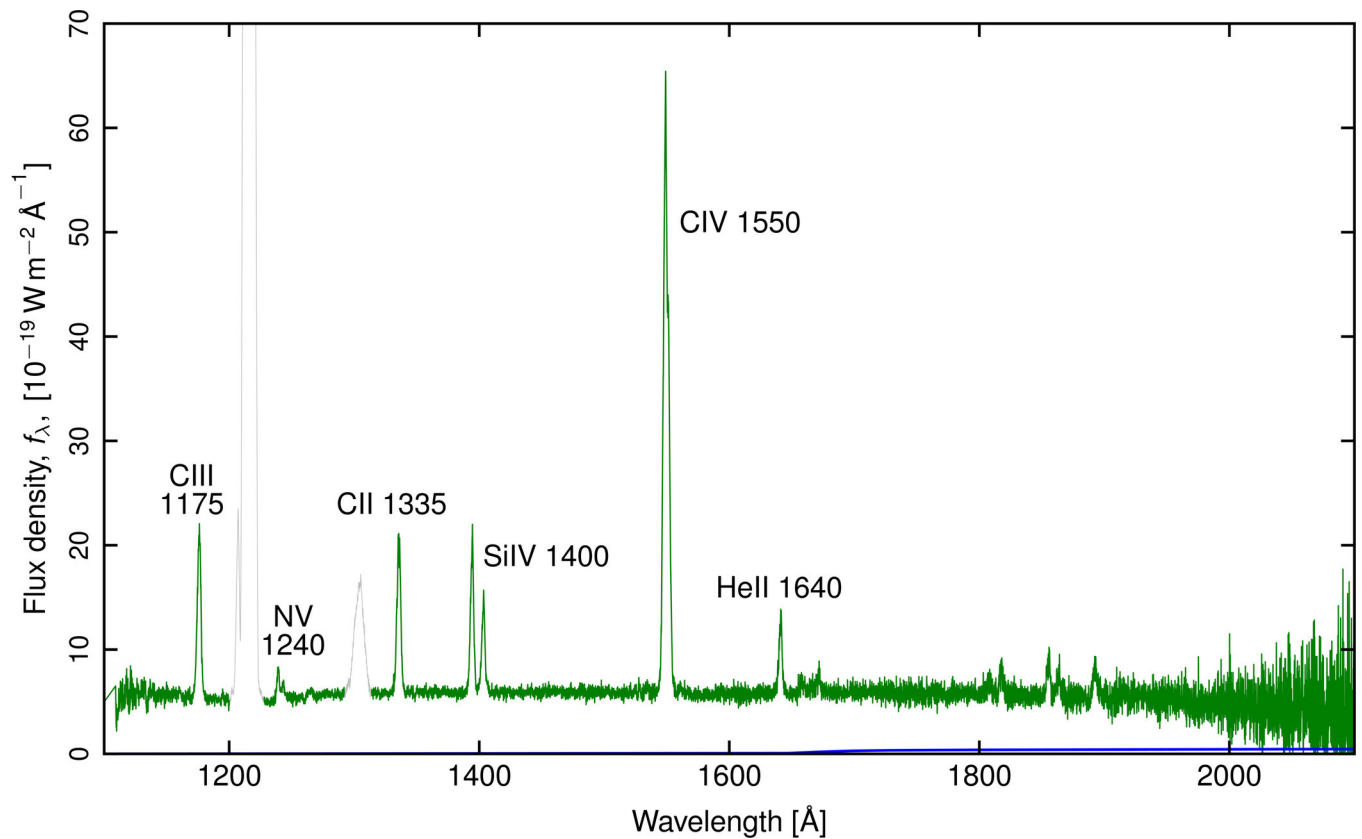
Code availability. The data were reduced with standard instrument pipelines for the HST, VLT, and Swift data. The WHT and INT data were reduced with STARLINK software. Scripts for creating the figures are available on request from T.R.M. The code for computing the white dwarf model atmosphere, which is a legacy F77 code and complex to export, is unavailable. The atmosphere model itself, however, is available on request from T.R.M.

30. Becker, R. H., White, R. L. & Helfand, D. J. The FIRST Survey: faint images of the radio sky at twenty centimeters. *Astrophys. J.* **450**, 559–577 (1995).
31. Skrutskie, M. F. *et al.* The Two Micron All Sky Survey (2MASS). *Astron. J.* **131**, 1163–1183 (2006).
32. Pilbratt, G. L. *et al.* Herschel Space Observatory—an ESA facility for far-infrared and submillimetre astronomy. *Astron. Astrophys.* **518**, L1–L6 (2010).
33. Wright, E. L. *et al.* The Wide-field Infrared Survey Explorer (WISE): mission description and initial on-orbit performance. *Astron. J.* **140**, 1868–1881 (2010).
34. Werner, M. W. *et al.* The Spitzer Space Telescope mission. *Astrophys. J. Suppl. Ser.* **154**, 1–9 (2004).
35. Murphy, T. *et al.* The Australia Telescope 20 GHz Survey: the source catalogue. *Mon. Not. R. Astron. Soc.* **402**, 2403–2423 (2010).
36. De Breuck, C., Tang, Y., de Bruyn, A. G., Röttgering, H. & van Breugel, W. A sample of ultra steep spectrum sources selected from the Westerbork In the Southern Hemisphere (WISH) survey. *Astron. Astrophys.* **394**, 59–69 (2002).
37. Dhillon, V. S. *et al.* ULTRASPEC: a high-speed imaging photometer on the 2.4-m Thai National Telescope. *Mon. Not. R. Astron. Soc.* **444**, 4009–4021 (2014).
38. Hambach, F.-J. ROAD (Remote Observatory Atacama Desert): intensive observations of variable stars. *J. Am. Assoc. Var. Star Obs.* **40**, 1003–1009 (2012).
39. Rebassa-Mansergas, A., Gänsicke, B. T., Rodríguez-Gil, P., Schreiber, M. R. & Koester, D. Post-common-envelope binaries from SDSS — I. 101 white dwarf main-sequence binaries with multiple Sloan Digital Sky Survey spectroscopy. *Mon. Not. R. Astron. Soc.* **382**, 1377–1393 (2007).
40. Hessman, F. V., Robinson, E. L., Nather, R. E. & Zhang, E.-H. Time-resolved spectroscopy of SS Cygni at minimum and maximum light. *Astrophys. J.* **286**, 747–759 (1984).
41. Wade, R. A. & Horne, K. The radial velocity curve and peculiar TiO distribution of the red secondary star in Z Chamaeleontis. *Astrophys. J.* **324**, 411–430 (1988).
42. Marsh, T. R. A spectroscopic study of the deeply eclipsing dwarf nova IP Peg. *Mon. Not. R. Astron. Soc.* **231**, 1117–1138 (1988).
43. Faulkner, J., Flannery, B. P. & Warner, B. Ultrashort-period binaries. II. HZ 29 (=AM CVn): a double-white semidetached postcataclysmic nova? *Astrophys. J.* **175**, L79–L83 (1972).
44. Knigge, C., Baraffe, I. & Patterson, J. The evolution of cataclysmic variables as revealed by their donor stars. *Astrophys. J. Suppl. Ser.* **194**, 28–75 (2011).
45. Kellermann, K. I. & Pauliny-Toth, I. I. K. The spectra of opaque radio sources. *Astrophys. J.* **155**, L71–L78 (1969).



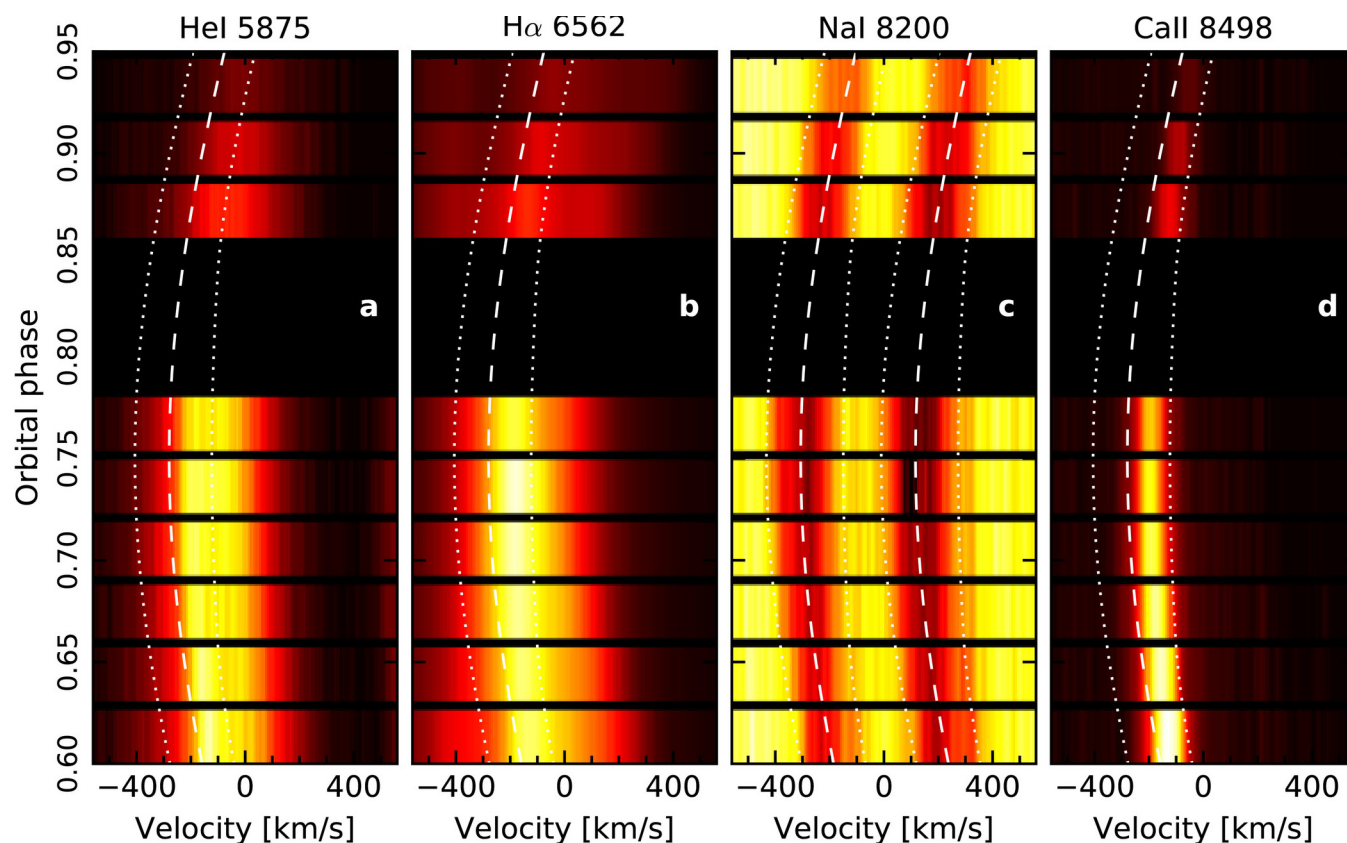
Extended Data Figure 1 | The optical spectrum of the white dwarf's M star companion. A 10 min exposure of AR Sco taken with FORS on the VLT between orbital phases 0.848 and 0.895 (black). Other spectra: an optimally scaled M5 template (green); the sum of the template plus a fitted smooth spectrum (red); AR Sco minus the template, that is, the extra light

(magenta); a white dwarf model atmosphere of $T = 9,750 \text{ K}$, $\log[g] = 8.0$, the maximum possible consistent with the HST data (blue). A slit-loss factor of 0.61 has been applied to the models. The strong emission lines come from the irradiated face of the M star.



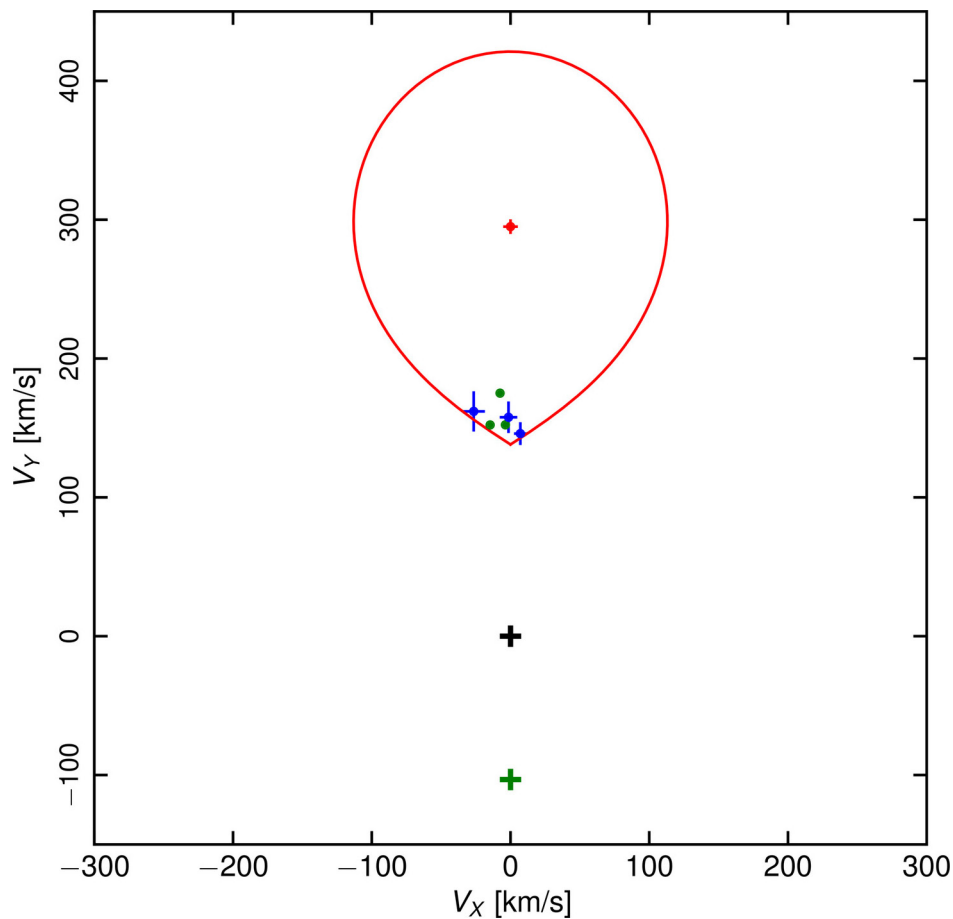
Extended Data Figure 2 | HST ultraviolet spectrum of AR Sco. This shows the mean HST spectrum with geocoronal emission plotted in grey. The blue line close to the x axis is a white dwarf model atmosphere of $T = 9,750 \text{ K}$, $\log[g] = 8.0$, representing the maximal contribution of

the white dwarf consistent with light curves. The radial velocities of the emission lines (Extended Data Fig. 4) show that, similar to the optical lines, the ultraviolet lines mainly come from the irradiated face of the M star.



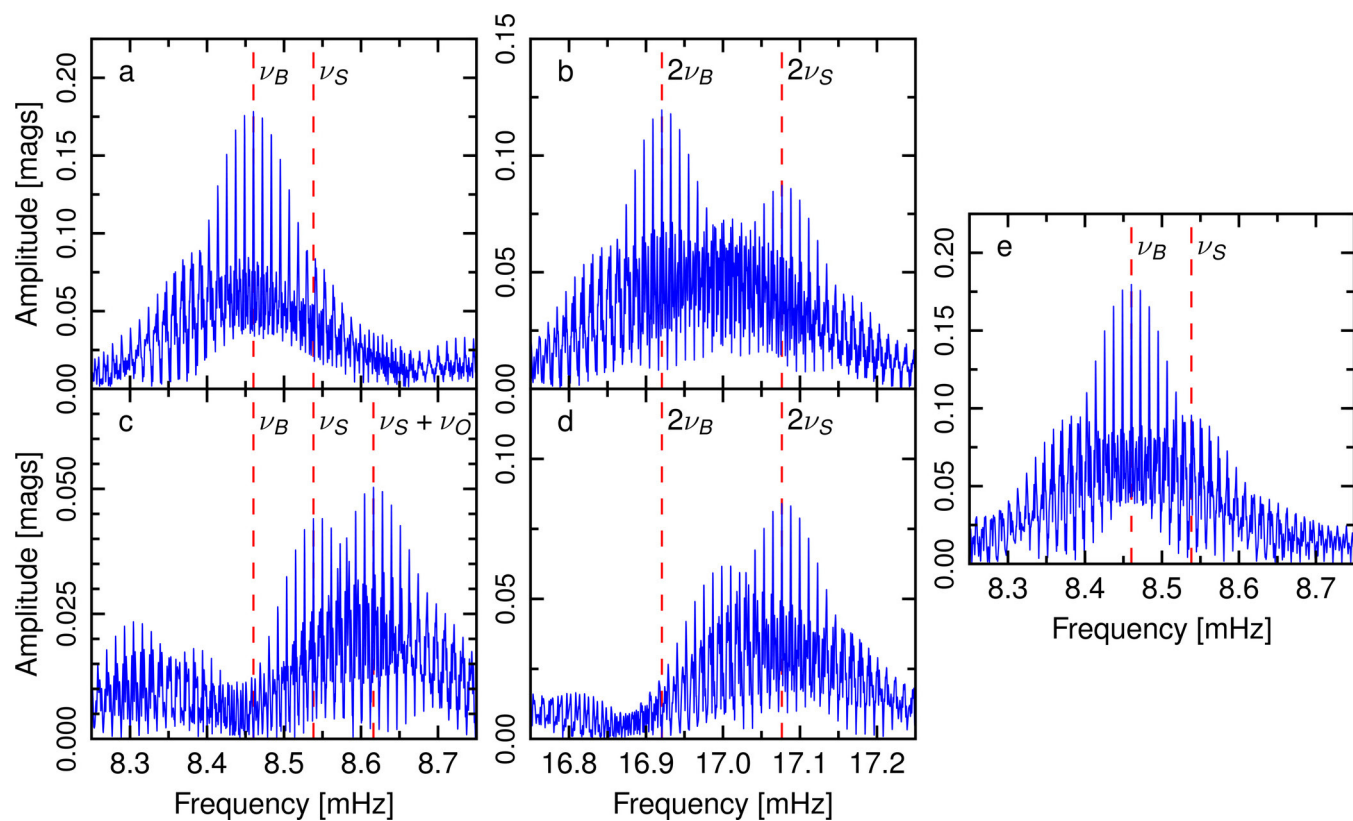
Extended Data Figure 3 | Velocity variations of atomic emission lines compared with those of the M star. a–d, Emission lines from a sequence of spectra from the VLT+X-SHOOTER data (a, b, d) and the Na I 8,200 absorption doublet from the M star (d). The dashed line shows the motion of the centre of mass of the M star deduced from the NaI measurements,

while the dotted lines show the maximum range of radial velocities from the M star for $q = M_2/M_1 = 0.35$. The emission lines move in phase with the Na I doublet but at lower amplitude, showing that they come from the inner face of the M star.



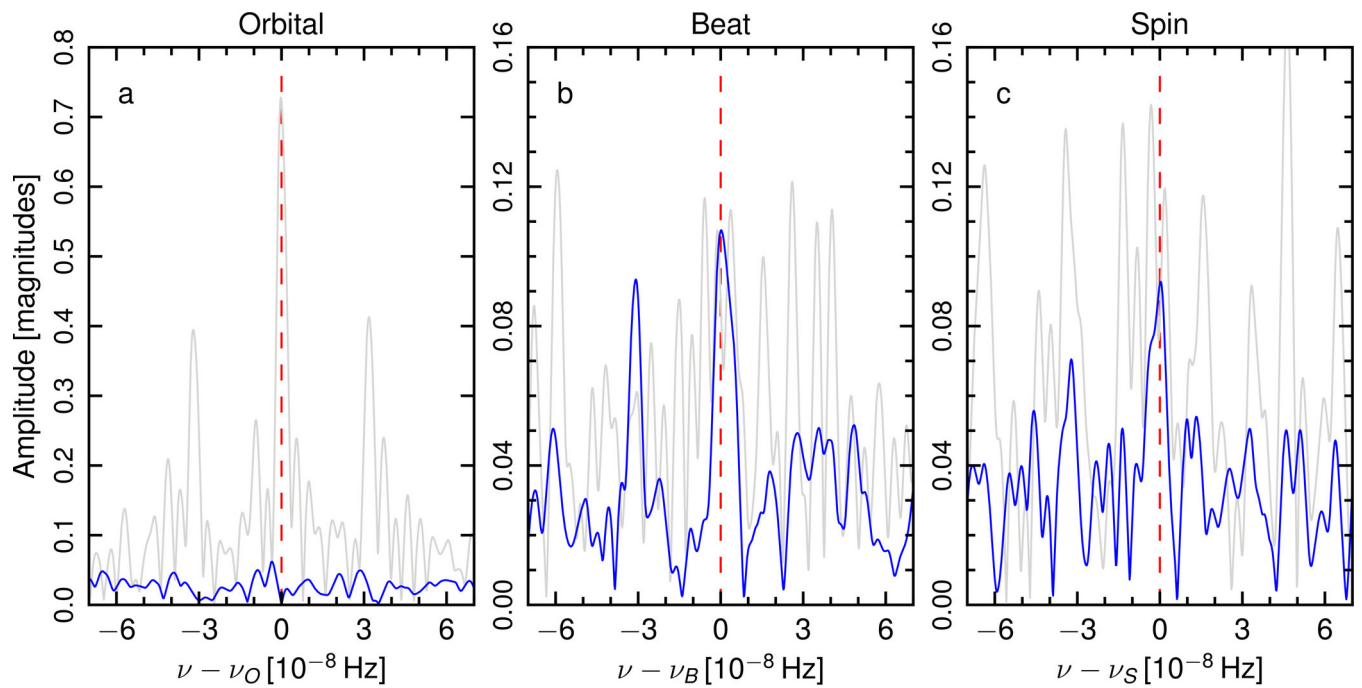
Extended Data Figure 4 | The origin of the emission lines relative to the M star. Velocities of the lines were fitted with $V_R = -V_X \cos(2\pi\varphi) + V_Y \sin(2\pi\varphi)$. The points show the values of (V_X, V_Y) . The M star from Na I is shown by the red dot (by definition this lies at $V_X = 0$). Si IV and He II lines from the HST FUV data are shown by the blue dots. H α , H β and H γ from optical spectroscopy are shown by the green dots. The black

and green plus signs mark the centres of mass of the binary and white dwarf, respectively. Error bars are $\pm 1\sigma$, calculated from fits to the radial velocities with uncertainties on the velocities scaled to result in $\chi^2 = 1$ per degree of freedom, and the uncertainties on the fit parameters calculated from the covariance matrix of the linear least-squares fit. The red line shows the Roche lobe of the M star for a mass ratio $q = 0.35$.



Extended Data Figure 5 | Amplitude spectra from nine days of monitoring with a small telescope. **a**, Amplitude as a function of frequency around the 1.97 min signal from data taken with a 40 cm telescope. **b**, The same at the second harmonic. **c**, **d**, The same as **a** and **b**

after subtracting the beat frequency signals at ν_B (**c**) and $2\nu_B$ (**d**). Signals at $\nu_S + \nu_O$ and $2\nu_S - \nu_O$ are also apparent. **e**, The window function, computed from a pure sinusoid of frequency ν_B and amplitude 0.18 magnitudes (see **a**).



Extended Data Figure 6 | Amplitude spectra from seven years of sparsely sampled CSS data. a–c, The amplitude as a function of frequency relative to the mean orbital (a), beat (b) and spin (c) frequencies listed in Extended Data Table 2. The grey line is the spectrum without any processing; the blue line is the spectrum after subtraction of the orbital signal.

Extended Data Table 1 | Observation log

Tel./Inst.	Type	Wavelength	Date	Exposure $T[s] \times N$
VLT+FORS	Spectra	420 – 900 nm	2015-06-03	600x1
WHT+ULTRACAM	Photometry	u', g', r'	2015-06-23	2.9x768
WHT+ULTRACAM	Photometry	u', g', i'	2015-06-24	1.3x7634
<i>Swift</i> +UVOT/XRT	UV, X-rays	260 nm, 0.2 – 10 keV	2015-06-23 – 2015-08-03	1000x10
VLT+HAWKI	Photometry	K_S	2015-07-06	2.0x7020
WHT+ISIS	Spectra	354 – 539, 617 – 884 nm	2015-07-16	20x94
WHT+ISIS	Spectra	354 – 539, 617 – 884 nm	2015-07-17	300x4
WHT+ISIS	Spectra	356 – 520, 540 – 697 nm	2015-07-19	30x130
ROAD 40 cm	Photometry	White light	2015-07-19 – 2015-07-28	30x1932
WHT+ISIS	Spectra	356 – 520, 540 – 697 nm	2015-07-20	30x210
INT+IDS	Spectra	440 – 685 nm	2015-07-22	27x300
INT+IDS	Spectra	440 – 685 nm	2015-07-23	34x300
ATCA	Radio	5.5, 9.0 GHz	2015-08-13	271x10
WHT+ISIS	Spectra	320 – 535, 738 – 906 nm	2015-08-26	600x8
WHT+ISIS	Spectra	320 – 535, 738 – 906 nm	2015-09-01	600x8
VLT+XSHOOTER	Spectra	302 – 2479 nm	2015-09-23	11x300
<i>HST</i> +COS	Spectra	110 – 220 nm	2016-01-19	5 orbits
TNT+ULTRASPEC	Photometry	g'	2016-01-19	3.8x1061

Extended Data Table 2 | Statistics of the orbital, beat and spin frequencies from bootstrap fits

Frequency	5%-ile mHz	95%-ile mHz	Median mHz	Mean mHz	RMS mHz
ν_{O}	0.077921311	0.077921449	0.077921380	0.077921380	0.000000042
ν_{B}	8.4603102	8.4603140	8.4603112	8.4603114	0.0000011
ν_{S}	8.5382332	8.5382356	8.5382348	8.5382346	0.0000008

Extended Data Table 3 | Archival data sources and flux values

Source	Wavelength, Frequency	Flux mJy	Source	Wavelength, Frequency	Flux mJy
WISH	352 MHz	< 18	WISE	22.0 μm	45.2 – 105.4
FIRST	1.4 GHz	8.0 ± 0.3	WISE	12 μm	18.0 – 48.3
AT20G	20 GHz	< 50	<i>Spitzer</i>	5.73 μm	11.9 – 23.5
<i>Herschel</i>	500 μm	92 ± 25	WISE	4.60 μm	11.8 – 20.5
<i>Herschel</i>	350 μm	76 ± 21	<i>Spitzer</i>	3.6 μm	13.0 ± 0.7
<i>Herschel</i>	250 μm	55 ± 23	WISE	3.4 μm	13.2 – 13.8
<i>Herschel</i>	160 μm	118 ± 38	2MASS	2.1 μm	13.5 ± 0.3
<i>Herschel</i>	70 μm	196 ± 63	2MASS	1.7 μm	15.0 ± 0.3
<i>Spitzer</i>	24 μm	59.9 ± 6.0	2MASS	1.2 μm	13.3 ± 0.3

Extreme creep resistance in a microstructurally stable nanocrystalline alloy

K. A. Darling¹, M. Rajagopalan², M. Komarasamy³, M. A. Bhatia², B. C. Hornbuckle¹, R. S. Mishra³ & K. N. Solanki²

Nanocrystalline metals, with a mean grain size of less than 100 nanometres, have greater room-temperature strength than their coarse-grained equivalents, in part owing to a large reduction in grain size¹. However, this high strength generally comes with substantial losses in other mechanical properties, such as creep resistance, which limits their practical utility; for example, creep rates in nanocrystalline copper are about four orders of magnitude higher than those in typical coarse-grained copper^{2,3}. The degradation of creep resistance in nanocrystalline materials is in part due to an increase in the volume fraction of grain boundaries, which lack long-range crystalline order and lead to processes such as diffusional creep, sliding and rotation³. Here we show that nanocrystalline copper–tantalum alloys possess an unprecedented combination of properties: high strength combined with extremely high-temperature creep resistance, while maintaining mechanical and thermal stability. Precursory work on this family of immiscible alloys has previously highlighted their thermo-mechanical stability and strength^{4,5}, which has motivated their study under more extreme conditions, such as creep. We find a steady-state creep rate of less than 10^{-6} per second—six to eight orders of magnitude lower than most nanocrystalline metals—at various temperatures between 0.5 and 0.64 times the melting temperature of the matrix (1,356 kelvin) under an applied stress ranging from 0.85 per cent to 1.2 per cent of the shear modulus. The unusual combination of properties in our nanocrystalline alloy is achieved via a processing route that creates distinct nanoclusters of atoms that pin grain boundaries within the alloy. This pinning improves the kinetic stability of the grains by increasing the energy barrier for grain-boundary sliding and rotation and by inhibiting grain coarsening, under extremely long-term creep conditions. Our processing approach should enable the development of microstructurally stable structural alloys with high strength and creep resistance for various high-temperature applications, including in the aerospace, naval, civilian infrastructure and energy sectors.

Over the past 50 years, the reduction or elimination of intrinsic topological defects (grain or cell boundaries) has been central to the design of creep-resistant materials. Current designs enhance high-temperature creep performance through the use of single-crystal alloys, for example, nickel-based, single-crystal superalloys^{6,7}. Nano-grained materials with grain sizes 7–8 orders of magnitude smaller than, and grain-boundary volume fractions 5–6 orders of magnitude higher than, the currently used single-crystal superalloys have never been considered viable for high-temperature creep applications. Moreover, nanocrystalline metals exhibit microstructural instability, that is, grain growth via diffusional processes such as diffusional creep, sliding and rotation, at moderately low and even room temperatures, sometimes in combination with deformation^{2,8,9}. Consequently, previous creep studies on nanocrystalline metals have reported creep-stress exponents of 1–3 resulting from grain-size effects on diffusional (Coble) creep².

Contrary to conventional wisdom, we have developed a divergent, bulk nanocrystalline copper–tantalum alloy (10 atomic per cent (at%) tantalum; hereafter Cu–10 at% Ta) that is able to achieve and retain high strength and creep resistance at a high homologous temperature of $0.64T_m \approx 600^\circ\text{C}$ (where T_m is the melting temperature of the matrix), owing to its unique microstructural architecture. Initially synthesized through high-energy ball milling and subsequently consolidated via equal-channel angular extrusion (ECAE), the as-processed microstructure consists of a copper matrix with an average grain size of 50 ± 17.5 nm and tantalum-based particles that range in size from atomic nanoclusters (average diameter of 3.18 ± 0.86 nm) to much larger precipitates (average diameter of 32 ± 7.5 nm); uncertainties given here and elsewhere are 1 s.d. (For a macroscopic view of the microstructure and additional processing details, see Extended Data Fig. 2.) It has been shown that such ranges in particle size give rise to an extremely stable microstructure^{4,5,10,11}; for example, as compared to pure nanocrystalline copper, which exhibits rapid grain growth to the micrometre scale at only 100°C (ref. 12), Cu–10 at% Ta powders maintain a mean grain size of 167 nm after annealing at $1,040^\circ\text{C}$ for 4 h (ref. 4). This highly stabilized microstructure could give rise to unusual combinations of mechanical properties, such as creep resistance under extreme conditions (high stress and temperature).

We conducted compression creep tests over a wide range of applied stress and temperature conditions (see Methods), as shown in Fig. 1. The compression–creep–strain evolution curves shown in Fig. 1a consist of the primary creep region, where the creep strain rate decreased with time, and the secondary creep region, where the creep strain rate remained at steady state. The steady-state creep rates in nanocrystalline Cu–10 at% Ta were all found to be less than 10^{-6} s^{-1} at various homologous temperatures between $0.5T_m$ and $0.64T_m$ under a stress of 1.2%–0.85% of the shear modulus. Note that the creep rates reported for nanocrystalline Cu–10 at% Ta are minimum creep rates. The upper and lower fractions of the shear modulus equate to stress values of 576 MPa and 319 MPa, respectively, which represent 90% and 65% of the at-temperature (400°C and 600°C) yield strength. The yield stress values at various temperatures were quantified using a series of quasi-static compression tests with a strain rate of $8 \times 10^{-4} \text{ s}^{-1}$; see Extended Data Fig. 4. To further demonstrate the improvement in the creep resistance, a creep test was performed under 100% of the yield stress at $0.5T_m$ (approximately 400°C), which resulted in a creep rate of $5.3 \times 10^{-8} \text{ s}^{-1}$. By contrast, at a rather low homologous temperature— $0.4T_m$ or about 275°C , for example—a creep rate of 10^{-1} s^{-1} was reported for an applied stress of 0.12% of the shear modulus (57 MPa) with an average grain size of 25 nm in pure nanocrystalline copper². Compared to pure nanocrystalline copper, nanocrystalline Cu–10 at% Ta at 1.5–2 times higher temperature and an order of magnitude higher stress has a creep rate that is 6–8 orders of magnitude lower (ref. 3). Such a response is reminiscent of, and more comparable to, that of the creep performance achieved by advanced single-crystal nickel-based superalloys (creep rate of about 10^{-8} s^{-1})¹³.

¹Army Research Laboratory, Aberdeen Proving Ground, Maryland 21005, USA. ²School of Engineering of Matter, Transport, and Energy, Arizona State University, Tempe, Arizona 85281, USA.

³Department of Materials Science and Engineering, University of North Texas, Denton, Texas 76203, USA.

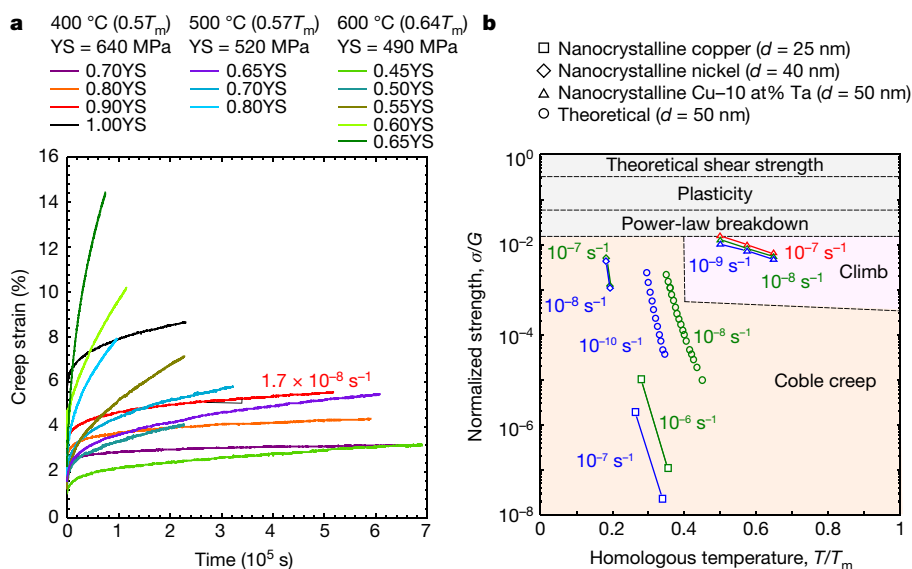


Figure 1 | Compressive creep response of nanocrystalline Cu-10 at% Ta. **a**, Conventional creep strain versus time for various applied temperatures and constant-stress conditions (given as fractions of the yield stress, YS), as indicated in the legend. For example, the creep rate for the red curve corresponds to a value of $1.7 \times 10^{-8} \text{ s}^{-1}$. **b**, Map of the theoretical deformation mechanisms of nanocrystalline copper with an average grain size of 50 nm (shaded regions), with theoretical, constant, Coble

creep rates for a grain size $d = 50$ nm (circles), experimental creep rates² in nanocrystalline copper (squares; $d = 25$ nm) and nickel (diamonds; $d = 40$ nm), and the creep rate we found for nanocrystalline Cu-10 at% Ta (triangles; $d = 50$ nm) over-plotted. The different coloured symbols correspond to different creep rates, as indicated in the legend. T , temperature; T_m , melting temperature of the matrix; σ , applied stress; G , shear modulus.

In general, creep in nanocrystalline materials has been reported to follow the Coble creep mechanism¹⁴, whereby creep occurs through the transport of vacancies along grain boundaries^{14,15} with a low stress exponent (of the order of 1–3)^{2,3}. On the other hand, our nanocrystalline Cu-10 at% Ta alloy exhibits stress exponents that are substantially higher than those associated with the diffusional-creep- and grain-boundary-related mechanisms (see Methods). Therefore, the

creep resistance achieved with our nanocrystalline Cu-10 at% Ta alloy outperforms most nanocrystalline materials. To comprehend this improvement in creep resistance, a compilation of experimental and theoretical creep-rate data for various nanocrystalline materials is presented on an Ashby-type deformation mechanism map¹⁵ in Fig. 1b, which was derived on the basis of creep constants for nanocrystalline copper with a mean grain size of 50 nm (see ref. 15 and Methods).

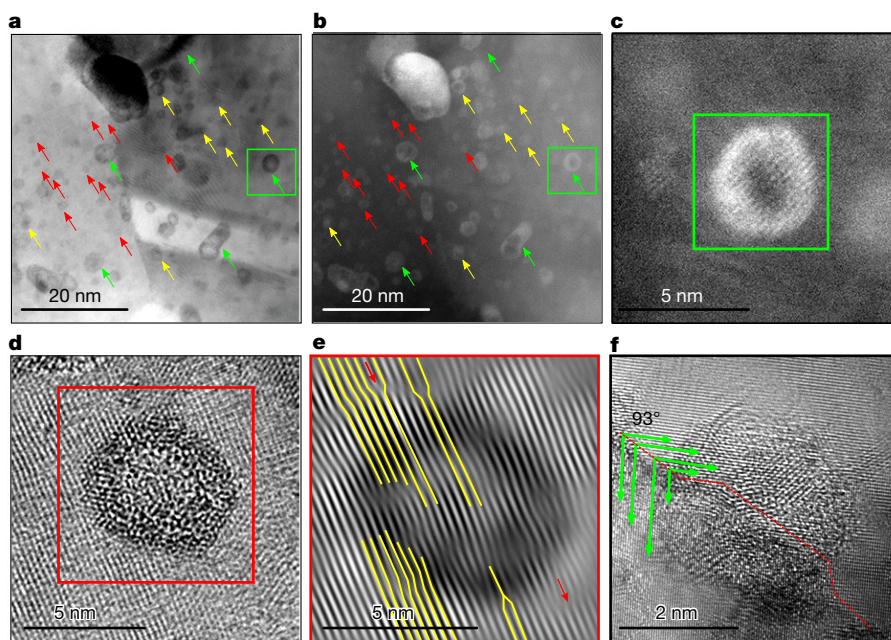


Figure 2 | TEM characterization of tantalum-based nanocluster in as-received nanocrystalline Cu-10 at% Ta. **a**, BF-STEM image highlighting the high density of nanoclusters of various sizes. The coloured arrows indicate the sizes (radii) of the different coherent or semi-coherent nanoclusters (red, approximately 1 nm; yellow, approximately 2.5 nm; green, 4 nm or greater). **b**, HAADF-STEM image accentuating tantalum-rich clusters on the basis of atomic-number contrast; arrows as in **a**. **c**, High-magnification HAADF image of the green-boxed area in

a and **b**. **d**, BF-STEM image highlighting the core-shell structure of the nanocluster. **e**, Inverse fast Fourier transform image of the red-boxed region in **d** highlighting the threading dislocations (yellow lines) and half planes (red arrows) between the matrix semi-coherent clusters. **f**, 3-nm-diameter particle residing at a high-angle (93°) grain boundary (red dashed lines). The green arrows correspond to the direction of lattice planes of the copper matrix.

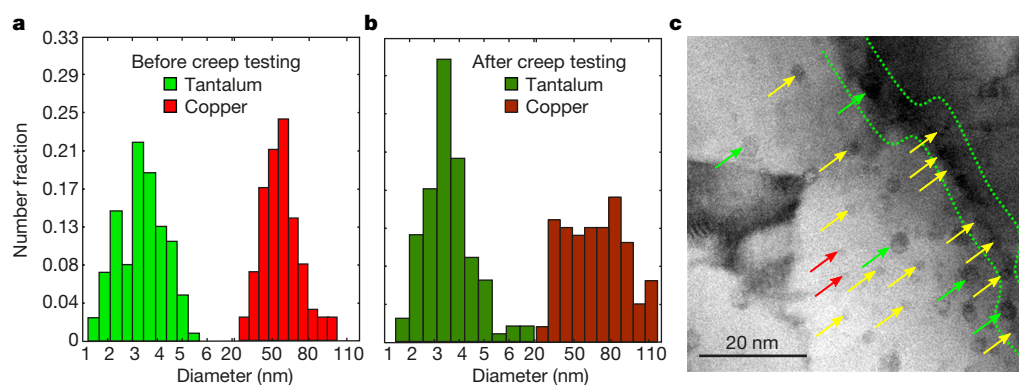


Figure 3 | TEM images showing the microstructures and grain distributions indicating stability in nanocrystalline Cu–10 at% Ta after creep testing. **a, b**, Number distributions determined from 300 grains of the copper matrix (red) and tantalum particles (green) before (**a**) and after (**b**) creep testing. The distributions of copper grains and tantalum

Experimental creep-rate data from nanocrystalline metals such as copper (25-nm grain size) and nickel (40-nm grain size)² along with the theoretical, constant, Coble creep-rate lines for copper with an average grain size of 50 nm (green and blue circles in Fig. 1b) are also presented. The reported creep properties of nanocrystalline copper and nickel fall within the Coble region. This is mainly due to grain-boundary diffusional processes; that is, vacancy diffusion and self-diffusion in copper and nickel, occurring through both the grain boundaries and the lattice are faster at elevated temperatures and, hence, the diffusional creep controls the creep behaviour³. Therefore, in these conventional nanocrystalline copper and nickel metals, the grain coarsening creates powerful kinetics that constantly evolves the microstructure. By contrast, the creep rates of our nanocrystalline Cu–10 at% Ta show a marked departure from convention, with the measured creep rates primarily in the dislocation–climb region (as shown by the triangles in Fig. 1b). In other words, the diffusional creep processes have been suppressed (or were absent) in our nanocrystalline Cu–10 at% Ta alloy.

To understand the observed enhancement of the creep property, we turn our attention to the large number (density of $6.5 \times 10^{23} \text{ m}^{-3}$) of coherent or semi-coherent (diameters of 1–4 nm) nanoclusters (see Fig. 2a, b). These small nanoclusters have a core–shell-type structure that can be seen in Fig. 2b, c, which shows that the contrast within and across the individual nanoclusters varies, indicating a compositional gradient. In addition, the high-angle annular dark-field scanning

particles before and after creep testing are similar, indicating little if any microstructural evolution. **c**, High-resolution BF-STEM image showing the bowing of the grain boundary (green dashed lines) as it interacts with tantalum clusters. The coloured arrows highlight the sizes (diameters) of the clusters (red, <1 nm; yellow, <1–2 nm; green, >4 nm).

transmission electron microscopy (HAADF-STEM) image (Fig. 2b) points to the shell portion of the particle being tantalum-rich, with the core generating less contrast, possibly owing to an element with a lower atomic number or to structural defects that would not generate contrast (such as vacancies). The high-resolution transmission electron microscopy (HR-TEM) image of the nanocluster (Fig. 2d) provides further evidence that the loss in contrast could be partly due to the presence of vacancies within the core region. The inverse fast Fourier transform image of one such nanocluster is shown in Fig. 2e. Note the distortion of the lattice as it approaches and enters the nanocluster, with the yellow arrows indicating the insertion of extra half planes of atoms into the lattice to minimize distortion. Finally, Fig. 2f shows a HR-TEM characterization of semi-coherent bonding between the nanocluster and the copper matrix at a high-angle (93°) grain boundary with an average misfit strain of 5.8%, indicating strong interfacial bonding that can lead to enhanced mechanical properties. Quasi-static and dynamic strengths of greater than 1.2 GPa have previously been measured for the nanocrystalline Cu–10 at% Ta alloy (see Methods); these strengths are greater than double that predicted by Hall–Petch hardening for nanocrystalline copper⁴ and presented with an apparent linear temperature dependence of flow stress⁵. Core–shell-type nanoclusters have recently been reported in oxide-dispersion strengthened (ODS) ferritic alloys¹⁶ and molybdenum alloys¹⁷, and are responsible for the excellent strength and ductility therein.

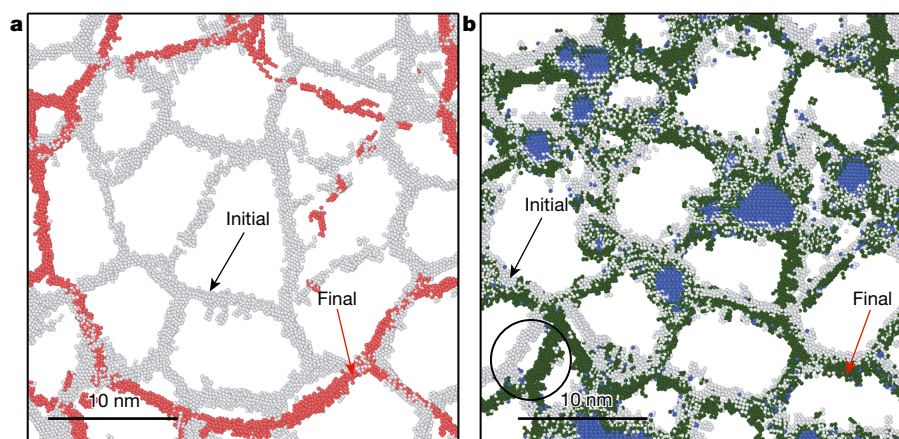


Figure 4 | Modelling data indicating stability in nanocrystalline Cu–10 at% Ta after creep testing. **a, b**, Two-dimensional slices through three-dimensional atomistic creep simulations of pure nanocrystalline copper (**a**) and nanocrystalline Cu–10 at% Ta (**b**) at 600 °C and 295 MPa of applied stress. White atoms represent the initial grain-boundary configurations (average grain size of 8 nm); red and green atoms represent

the extent of coarsening associated with plastic deformation under constant load and temperature conditions. In **b**, tantalum atoms (blue) formulate a random distribution of grain-boundary clusters and localized growth is observed (circled in black) owing to insufficient Zener pinning in some grains.

To understand the underlying mechanisms of creep resistance and to determine the enhancement of the creep property induced by the nanoclusters, atomistic simulations and post-HR-TEM characterizations were performed (see Figs 3 and 4). First, the HR-TEM characterizations of post-deformed creep samples (at 600 °C and 50% yield stress) were performed, as shown in Fig. 3. The stability of the nanoclusters, which is crucial for enhanced properties and can be seen in Fig. 3a, b, is indicated by the coarsening rate of nanoclusters during creep at elevated temperatures being negligible, mainly owing to the coherency of such dispersions. Further, owing to highly stabilized nanoclusters, bowing of the grain boundary as it interacts with numerous nanoclusters can clearly be identified in a high-resolution bright-field STEM (BF-STEM) image of the post-creep sample, as shown in Fig. 3c. The bowing indicates that clusters located at grain boundaries are likely to increase the energy barrier for grain-boundary sliding and rotation, both of which are crucial creep mechanisms in nanocrystalline metals. In addition, such nanoclusters pin the grain boundaries (Zener pinning¹⁸), thereby preventing substantial grain coarsening, consistent with the atomistic calculations (Fig. 4a, b). Here, atomistic simulations were performed using the molecular-dynamics code LAMMPS¹⁹ and an embedded atom potential²⁰ (see Methods). Therefore, highly stabilized nanoclusters with strong structural affinity within the matrix and along the grain boundary are the microstructural features that governing the unusual combination of materials properties—high strength, extreme thermal stability and creep resistance.

Our results will lead to advances in designing nanocrystalline alloys with many simultaneously enhanced high-temperature properties, similar to those exhibited by creep-resistant single crystals, but with the additional benefit of much higher strength. For example, we show that a steady-state creep rate of less than 10^{-6} s^{-1} is attained at even $0.64T_m$ under a high applied stress (1.2% of the shear modulus). The creep rates in nanocrystalline Cu–10 at% Ta reported here are 6–8 orders of magnitude lower than most of the previously reported creep rates in nanocrystalline metals. We expect that the divergent creep behaviour reported here will change the theoretical understanding and expectations of the ways in which nanocrystalline metals deform at high temperatures and will result in new applications and capabilities.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 May; accepted 8 July 2016.

1. Gleiter, H. Nanostructured materials: basic concepts and microstructure. *Acta Mater.* **48**, 1–29 (2000).
2. Mohamed, F. A. & Li, Y. Creep and superplasticity in nanocrystalline materials: current understanding and future prospects. *Mater. Sci. Eng. A* **298**, 1–15 (2001).
3. Chokshi, A. H. Unusual stress and grain size dependence for creep in nanocrystalline materials. *Scr. Mater.* **61**, 96–99 (2009).

4. Darling, K. A. *et al.* Microstructure and mechanical properties of bulk nanostructured Cu–Ta alloys consolidated by equal channel angular extrusion. *Acta Mater.* **76**, 168–185 (2014).
5. Darling, K. A., Huskins, E. L., Schuster, B. E., Wei, Q. & Kecskes, L. J. Mechanical properties of a high strength Cu–Ta composite at elevated temperature. *Mater. Sci. Eng. A* **638**, 322–328 (2015).
6. Giamei, A. F. Development of single crystal superalloys: a brief history. *Adv. Mater. Process.* **171**, 26–30 (2013).
7. Reed, R. C. *The Superalloys: Fundamentals and Applications* (Cambridge Univ. Press, 2008).
8. Meyers, M. A., Mishra, A. & Benson, D. J. Mechanical properties of nanocrystalline materials. *Prog. Mater. Sci.* **51**, 427–556 (2006).
9. Choi, I.-C. *et al.* Nanoscale room temperature creep of nanocrystalline nickel pillars at low stresses. *Int. J. Plast.* **41**, 53–64 (2013).
10. Hornbuckle, B. C. *et al.* Effect of Ta solute concentration on the microstructural evolution in immiscible Cu–Ta Alloys. *JOM* **67**, 2802–2809 (2015).
11. Bhatia, M., Rajagopalan, M., Darling, K. A., Tschopp, M. A. & Solanki, K. N. The role of Ta on twinnability in nanocrystalline Cu–Ta alloys. *Mater. Res. Lett.* <http://dx.doi.org/10.1080/21663831.2016.1201160> (2016).
12. Huang, Y., Menovsky, A. & De Boer, F. Calorimetric analysis of the grain growth in nanocrystalline copper samples. *Nanostruct. Mater.* **2**, 587–595 (1993).
13. Pollock, T. M. & Tin, S. Nickel-based superalloys for advanced turbine engines: chemistry, microstructure and properties. *J. Propuls. Power* **22**, 361–374 (2006).
14. Coble, R. A model for boundary diffusion controlled creep in polycrystalline materials. *J. Appl. Phys.* **34**, 1679–1682 (1963).
15. Ashby, M. F. A first report on deformation-mechanism maps. *Acta Metall.* **20**, 887–897 (1972).
16. Hirata, A. *et al.* Atomic structure of nanoclusters in oxide-dispersion-strengthened steels. *Nat. Mater.* **10**, 922–926 (2011).
17. Liu, G. *et al.* Nanostructured high-strength molybdenum alloys with unprecedented tensile ductility. *Nat. Mater.* **12**, 344–350 (2013).
18. Manohar, P., Ferry, M. & Chandra, T. Five decades of the Zener equation. *ISIJ Int.* **38**, 913–924 (1998).
19. Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).
20. Purja Pun, G. P., Darling, K. A., Kecskes, L. J. & Mishin, Y. Angular-dependent interatomic potential for the Cu–Ta system and its application to structural stability of nano-crystalline alloys. *Acta Mater.* **100**, 377–391 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements M.R., and K.N.S. acknowledge the use of facilities within the LeRoy Eyring Center for Solid State Science at Arizona State University. This work was supported by US Army Research Laboratory under contract W911NF-15-2-0038. K.A.D. acknowledges A. J. Roberts and T. Luckenbaugh for synthesis of the Cu–Ta powder.

Author Contributions K.A.D., K.N.S. and R.S.M. equally contributed to the idea. K.A.D. and B.C.H. processed the nanocrystalline materials. M.K. performed the creep experiments. M.R. characterized the microstructure data and analysed deformation-mechanisms maps. M.A.B. performed the modelling work. K.N.S., K.A.D., M.R. and R.S.M. analysed the data. K.N.S., K.A.D., M.R. and R.S.M. wrote the paper. K.A.D. and M.R. edited the figures. K.A.D. supervised B.C.H., R.S.M. supervised M.K., and K.N.S. supervised M.R. and M.A.B.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.N.S. (kiran.solanki@asu.edu).

Reviewer Information Nature thanks J. Cormier, S. Forest and T. Perez Prado for their contribution to the peer review of this work.

METHODS

Powder processing and consolidation via ECAP. For the preparation of nanocrystalline (NC) Cu–10 at % Ta powder, the powder was generated through high-energy cryogenic mechanical alloying. The desired composition was obtained by loading elemental Cu and Ta powders (–325 mesh and 99.9% purity) into a hardened steel vial along with the milling media (440C stainless steel balls) inside a glove box with an Ar atmosphere (oxygen and H₂O are <1 p.p.m.). The vials were loaded with 10 g of the Cu–Ta powder as well as the appropriate amount of media to ensure a ball-to-powder ratio of 5:1 by weight. A SPEX 8000 M shaker mill was used to perform the milling at cryogenic temperature (verified to be about –196 °C) for 4 h (14.4 ks) using liquid nitrogen. To ensure the vial remained at cryogenic temperature, a thick polymer sleeve was retrofitted to fit around the vial in the SPEX mill with an inlet and outlet vent to flow the liquid nitrogen. Before starting the milling process, the vial was placed in the polymer sleeve with the liquid nitrogen flowing for approximately 20 min (1.2 ks) to ensure the vial approached –196 °C. Once the milling was completed, the vials were placed back into the glove box, opened and stored. This milling procedure was performed until 100 g of NC Cu–10 at % Ta powder was generated. The resulting powder after cryogenic mechanical milling was an unagglomerated mass of powder with particulates ranging in size from about 20 µm to 100 µm.

For consolidating the NC Cu–10 at % Ta powder to bulk, equal-channel angular extrusion (ECAP) was selected as the consolidation process. Billets of Ni 201 with dimensions of 25.4 mm × 25.4 mm × 90 mm had cylindrical chambers with diameters of 10 mm and lengths of 50 mm made within them for housing the powder. The powder was loaded into the chamber followed by press-fitting a Ni 201 plug into the open end to seal the chamber. Both of these steps were performed within the glove box. Before starting the ECAP process, the die assembly used for processing the billets was preheated to 350 °C to minimize thermal loss during the ECAP processing. Additionally, the billets containing the powder were held at 700 °C in a box furnace purged with Ar for 40 min (2.4 ks) to ensure that they reach the desired extrusion temperature. The heated billets were dropped into the ECAP tooling as quickly as possible from the furnace and extruded at an extrusion rate of 25.5 mm s^{–1}. This step was repeated four times following route 'BC'^{21–23} to prevent imparting a texture to the consolidated powder. As a result of the extrusion channel having an angle of 90°, a total strain of 460% was imparted onto the powder-containing billet as a result of processing. The creep specimens were then machined from these billets, within the region containing the consolidated powder, via a wire electric discharge machine. Finally, SEM imaging confirmed the creep specimens to be fully consolidated after the ECAP process with no porosity or as-milled particle boundaries being present. Note that a change in processing conditions or steps, such as in ECAP process temperatures, will result in different microstructural statistics such as grain-size distributions; however, as shown previously, the nanocluster density depends mainly on the Ta concentrations, which are the primary features resulting in an enhanced creep behaviour¹⁰.

Impurity levels. Impurities are a concern for all material processing techniques, including mechanical alloying via ball milling. During ball milling, the powder can pick-up impurities as a result of being exposed to the atmosphere and from the milling media themselves. To minimize O contamination, all powders were stored (before and after processing) and loaded into vials and billets under an Ar atmosphere (O and H₂O are <1 p.p.m.) inside a glove box. Additionally, to reduce the level of Fe contamination, the milling vial and bearings were coated in Cu by pre-milling the vial and the required bearings with high-purity Cu powder for 2 h at cryogenic temperatures. This process was repeated several times and produced a smooth, even coating of Cu over all milling surfaces (that is, the interior vial and exterior bearing surfaces). Despite these steps, energy-dispersive X-ray spectroscopy (EDS) analysis detected approximately 0.75 at % O in the bulk of the alloy¹⁰. To verify this O level, atom probe tomography (APT) was performed on as-milled powder and as-milled powder that was annealed for 1 h at 450 °C under a reducing atmosphere and NC Cu–10 at % Ta ECAP processed at 700 °C. APT results found all conditions to contain less than 1.25 at % O (ref. 10). Consequently, the O contamination in the alloy was minimized by following the procedural steps highlighted earlier. Finally, Fe contamination from the milling media was also detected via EDS, but could not be accurately measured; therefore, atom probe tomography was used again. From the APT analysis, the Fe contamination was found to vary between atom probe tips; however, the highest Fe content found was 1 at % and the lowest was 0.05 at % (ref. 10). This range indicates that the contamination from Fe is also minimal.

Microstructural characterization. X-ray diffraction was performed on as-received samples using an X'Pert PRO PANalytical MPD X-ray diffractometer with a Cu K α ($\lambda = 0.1542$ nm) radiation source. Owing to the resolution limit, the grain-size estimates from Scherrer's equation for the Cu matrix and Ta phase were inaccurate. Extended Data Fig. 1 indicates the X-ray reflections from Cu and Ta for the as-received condition, for which a random texture can be identified. Therefore,

to quantify the grain sizes and microstructure, transmission electron microscopy (TEM) was used. TEM characterizations were carried out in the as-received and post-deformed conditions using aberration-corrected ARM-200F and JEOL-2010F at 200 kV. Several images were captured in bright-field and high-resolution TEM as well as STEM mode to analyse the microstructure and quantify the statistics such as the grain-size distribution. The TEM samples were prepared through conventional thinning procedures whereby a 3-mm disk from the bulk specimen was thinned to about 70 µm and then dimpled to a thickness of about 5 µm. Ion milling was performed under liquid nitrogen temperatures to obtain electron-transparent regions in the specimens. The samples were also plasma cleaned in Ar before TEM observations to minimize contamination.

As-received microstructure. The primary microstructural characterization (Extended Data Fig. 2) of the as-received NC Cu–10 at % Ta ECAP processed at 700 °C revealed the presence of binary phases of Cu and Ta consistent with the X-ray diffraction measurements. The TEM characterization and precession diffraction data are illustrated in Extended Data Fig. 2, and demonstrate a high degree of randomness in the orientation relationship between the grains of NC Cu matrix with an average equiaxed grain diameter of 50 ± 17.5 nm. Orientation details were extracted from a region in the sample using the TOPSPIN software (resolution of 2 nm) on the TEM where a precession diffraction technique was used. In this technique, the incident electron beam is tilted and precessed along a conical surface, having a common axis with the TEM optical axis²⁴. Even though our NC material was consolidated to bulk, through severe plastic deformation, at 700 °C with a total accumulated strain of 4.6 (460%), the as-received averaged grain sizes were still in a NC regime (Extended Data Fig. 2). Further, Extended Data Fig. 2 shows a histogram, taken from multiple images similar to Extended Data Fig. 2a, b, indicating that the Ta particle size distribution has an average diameter of 32 ± 7.5 nm. A lower-magnification bright-field TEM image (Extended Data Fig. 2c) provides more insight into the much smaller Ta-based particles (diameters of <32 nm) and the presence of nano-twins within the NC Cu grains. Twinning is another important deformation mechanism in NC Cu that can be suppressed by the presence of fine nanoclusters¹¹. Further, the processing route produces a wide range of Ta particle sizes, ranging from atomic nanoclusters (average diameter of 3.18 ± 0.86 nm) to much larger precipitates (see Fig. 3 and Extended Data Fig. 2d). The energy of the interface²⁵ between the nanoclusters and the Cu matrix can be used to quantify the type of coherency and the cluster diameters over which the degree of coherency persists. Characterizing the coherency has indicated that this material has coherent, semi-coherent and incoherent nanoclusters (diameters of <3.898 nm, 3.898–15.592 nm and >15.592 nm, respectively). The nanoclusters also have misfit lattice dislocations at the interface that are indicative of the misfit strain present, which was identified using inverse fast Fourier transform analysis (Extended Data Fig. 3). On average, the misfit strain in the as-received sample is about 5.8%, but can be as high as 11%.

Mechanical characterization at quasi-static conditions. Quasi-static compression and tension tests of specimens over a temperature range from ambient up to 1,000 °C were performed using an Instron load frame equipped with a 10 kN and a 50 kN load cell, respectively, and an Applied Test Systems (ATS) clam-shell heating furnace capable of reaching maximum temperature of 1,500 °C. The specimens for compression were cylinders of 3 mm in diameter and length (aspect ratio of 1.0), whereas rectangular dogbones with lengths, widths and thicknesses of 3 mm, 1 mm and 1 mm, respectively, were used for tension. Tests were conducted at 24 °C, 200 °C, 400 °C, 600 °C, 800 °C, 900 °C and 1,000 °C with a strain rate of 8×10^{-4} s^{–1} for compression⁵ and 1×10^{-3} s^{–1} for tension. The system was held at the testing temperature for 15 min before loading to provide uniform temperature within the specimen. The push rods of the load frame were constructed out of precision-machined ZrO₂ rods to minimize heat losses. Boron nitride lubricated, polished tungsten carbide (WC) disks were used as platens for compression testing. Specimens were loaded under displacement control with an across head displacement of 0.15 mm min^{–1}. The force–displacement data were compliance corrected for all tests. The stress–strain responses are provided in Extended Data Fig. 4. The compressive curves in Extended Data Fig. 4a exhibit behaviour from elastic to nearly perfectly plastic over the entire temperature with no substantial strain hardening. Furthermore, the flow stress presented with an apparent linear temperature dependence^{4,5} as compared to the expected sigmoidal manifestation expected for pure coarse-grained Cu. Moreover, the Cu grain size after testing at 800 °C was estimated to be about 90 nm, indicating that grain coarsening is very limited and the reduction in observed yield and flow stress is a result of increased thermal softening only⁵. Therefore, NC Cu–10 at % Ta exhibits an extremely stable microstructure and unusual mechanical properties. In general, face-centred cubic materials such as Cu should not show any tension–compression asymmetry, which is evident from Extended Data Fig. 4b. The response in tension is perfectly elastic–plastic in nature with negligible strain hardening, identical to the compression tests. This response has implications for the tensile creep behaviour, where

this material will be expected to behave in a similar way for tensile-type creep tests as for compression.

Mechanical characterization at creep conditions. Compressive cylindrical creep experiments were performed using a 2320 series lever arm creep tester (Applied test systems) with a 5:1 lever arm ratio. Both the diameter and the height of the cylindrical creep specimens were about 3 mm. The specimens were kept at the centre of a 3210 series split tube furnace to maintain a constant temperature across the sample height. A heating rate of $200\text{ }^{\circ}\text{C h}^{-1}$ and a soak time of 0.5 h were used for the creep tests. For the best temperature measurement and control, a thermocouple was always wrapped around the creep specimens to maintain good contact. An ST 1278 incremental length gauge with $\pm 1\text{ }\mu\text{m}$ accuracy was used to measure the conventional creep strain. The compression creep experiments were conducted in air at 873 K and with fractions of 0.45, 0.50, 0.55, 0.60 and 0.65 of the yield stress, at 773 K and with fractions of 0.70, 0.75 and 0.80 of the yield stress, and at 673 K and with fractions of 0.70, 0.80, 0.90 and 1.00 of the yield stress. The specimens were first coated with a thin layer of boron nitride for lubrication and then placed between the compression platens. Creep test temperatures were attained at a constant heating rate followed by soaking at the set temperature (for 0.5 h) to avoid the temperature fluctuation during the test. After the soaking stage, the loading begins automatically, followed by the start of the creep test. These tests were typical constant-force tests. All the creep data were recorded from the test start to finish. Further, specimens did not reach failure because tests were stopped before the strain rate exponentially increases with stress (before the tertiary creep domain), and our primary objective was to characterize the secondary creep rates. For most of the creep tests, the total strain values did not exceed about 6%. All the crept samples were quenched in water immediately after unloading to preserve the crept microstructure. The physical dimensions of the crept samples were measured after the test and compared with the extensometer measurements. The stress was determined after the test by taking into account the amount of strain. Further, during loading the initial strain of the creep test specimen can include both elastic and plastic strains. The minimum creep rate was calculated from the slope of the curve of conventional creep strain versus time.

Theoretical deformation map. Theoretical deformation maps identify the deformation modes by which a polycrystalline material can deform¹⁵. In the case of creep deformation maps, the dominant mechanism is defined by considering the stress and temperature values for a particular value of the steady-state creep rate. The upper bound, that is, the theoretical shear strength, depicts the limit beyond which flow is possible even in defect-free crystals. This value of stress is of the order of the shear modulus and is independent of temperature. In the case of materials with defects, the motion of dislocations contributes toward plastic deformation; that is, the dislocation mechanisms are glide, climb and temperature-dependent dislocation creep. In the case of dislocation glide creep, impurities, solutes, precipitates and so on that are present in the material provide obstacles to plastic flow. At high temperatures, the dislocation creep mechanism is predominant where the deformation is controlled by diffusion and the strain rate is a nonlinear function of stress. Further, the motion of point defects leads to plastic deformation through either the grains (Nabarro–Herring^{26,27}) or the grain boundaries (Coble)¹⁴. These diffusional processes are independent of each other and depend on only the temperature¹⁵. The mechanisms that relate the steady-state creep rate $\dot{\epsilon}$ to the applied stress can be depicted using the equation¹⁵

$$\dot{\epsilon} = AD_0 \exp\left(-\frac{Q}{RT}\right) \left(\frac{Gb}{kT}\right) \left(\frac{b}{d}\right)^p \left(\frac{\sigma}{G}\right)^n$$

where A is a dimensionless constant, D_0 is a frequency factor, Q is the activation energy, R is the gas constant, T is the temperature, G is the shear modulus at the particular temperature, b is the burgers vector, k is the Boltzmann constant, d is the grain size, p is the grain-size exponent, n is the stress exponent and σ is the applied stress. The value of the constant A and exponents p and n depend on the mechanism considered. The values for the constants can be found in ref. 15. After incorporating the threshold stress, the rate-controlling creep deformation mechanism in the high stress and temperature regime was identified from the deformation map (Extended Data Fig. 5d) to be dislocation climb, where the apparent stress exponents of 10–18 were reduced to 4–8 (true stress exponents).

Creep rate data as a function of applied load with and without threshold correction. NC Cu–10 at% Ta processed at 700 $^{\circ}\text{C}$ and subjected to creep exhibits a high stress exponent n , as evident from the plot in Extended Data Fig. 5. To rationalize the high n values, appropriate threshold stress values were determined using a standard linear extrapolation method²⁸ whereby creep rate curves were plotted as a function of applied instantaneous stress ($\dot{\epsilon}^{1/n}$ versus σ) at various temperatures (Extended Data Fig. 5b). The data points fit onto a straight line that, on extrapolation to zero strain rates, yields a threshold value. The inadequacy with this method is that multiple straight lines exist for different n values provided the

experimental data covers a range of strain rates. The stress exponent value of five for threshold correction was deemed appropriate for this study, and corresponds to a dislocation-climb-based deformation mechanism whereby the threshold stress arises owing to the influence of dislocations in the creep process. Using the approximation, a threshold stress of 165 MPa was deduced and subtracted from the applied stress to illustrate the relation between normalized stress and creep rate (Extended Data Fig. 5c). The stress exponent values were computed to be between four and eight at various temperatures after threshold correction, indicating the absence of diffusional creep processes during creep. The data obtained through threshold correction for NC Cu–10 at% Ta and literature data² for NC Cu and NC Ni are plotted in Extended Data Fig. 5d. It is evident that NC Cu–10 at% Ta exhibits extreme creep resistance, with an increase in stress by an order of magnitude resulting in a 6–8-fold decrease in $\dot{\epsilon}$.

Atomistic modelling. The qualitative atomistic simulations were performed using a large-scale atomic/molecular massively parallel simulator (LAMMPS)¹⁹ along with a semi-empirical embedded atom potential (EAM) reported in ref. 20. This EAM potential was parameterized using an extensive database of energies and configurations from density functional theory (DFT) calculations of energy differences between various crystal structures of pure Cu and pure Ta, the formation energies of coherent Cu–Ta interfaces, and the binding energy of several ordered compounds, such as $\text{Li}_2\text{–Cu}_3\text{Ta}$, $\text{Li}_0\text{–CuTa}$, $\text{Li}_1\text{–CuTa}$, $\text{B}_2\text{–CuTa}$ and $\text{Li}_2\text{–Ta}_3\text{Cu}$ (ref. 20). See ref. 20 for more details on the validation of the EAM potential at different temperatures. The Voronoi tessellation method²⁹ was used to construct 3D NC Cu with an average grain size of 8 nm. Further, in the same NC Cu sample, spherical Ta particles with random distribution and size (average sphere radius of 0.7 nm) were doped to obtain the 10% Ta concentrations. The total number of atoms in a simulation cell was 1.4 million (with approximate box sizes of $35\text{ nm} \times 38\text{ nm} \times 15\text{ nm}$). The samples were first relaxed at the desired temperature using an NVT (conserving the number of atoms, volume and temperature) ensemble for 5 ns, followed by an independent relaxation in three directions using an NPT (conserving the number of atoms, pressure and temperature to mimic bulk behaviour) ensemble for another 5 ns with zero pressure in all the directions. These relaxations were performed to uniformly distribute the excess free energy through the whole system. Atomistic simulations were carried out using a molecular dynamics time step of 1 fs. Periodic boundary conditions were adopted in all directions. Then, the samples were loaded under tension along the y axis with a strain rate of 10^8 s^{-1} and at 600 $^{\circ}\text{C}$ while maintaining periodic and pressure-free boundary conditions along the x and z directions, respectively. The tensile simulations (up to 3% strain) were performed to increase the defect (dislocation twin, stacking fault) densities before the creep simulation to mimic the as-received experimental sample microstructure, as seen in Extended Data Fig. 6.

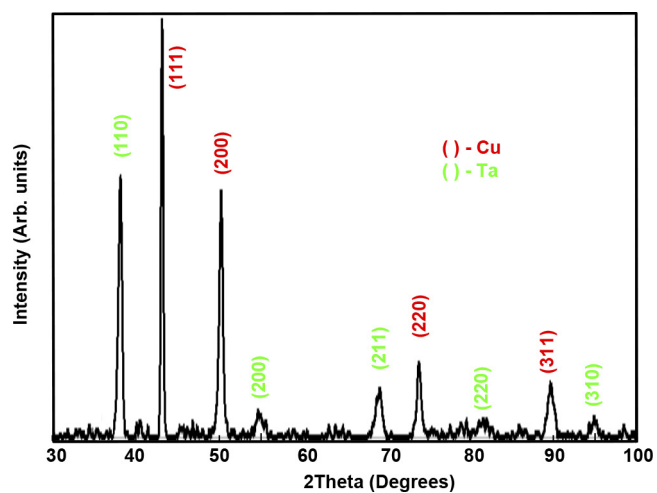
Finally, the NC and NC Cu–10 at% Ta simulation models were crept at 600 $^{\circ}\text{C}$ and 295 MPa applied stress along the y direction, whereas deformation in the other two directions was carried out by maintaining a zero pressure. The desired stresses were applied in the incremental form (a 5-MPa step) until 295 MPa, and then simulations were run at a constant applied stress for 5 ns or until failure. To overcome the relatively short time interval of the molecular dynamics simulation, we performed simulations at elevated temperatures, at which the distinct effects of the resultant liquid-like, fast grain-boundary diffusion (such as grain growth and microstructural instability), if present, were clearly identifiable (see refs 30, 31). Extended Data Fig. 7 shows minimum topological changes with the addition of Ta. Grain growth typically observed in NC materials due to both the stress-induced grain-boundary diffusive fluxes and grain-boundary sliding is hindered in Cu–Ta alloys³. The creep simulation of NC Cu and NC Cu–10 at% Ta in shown in Supplementary Video 1.

Microstructural stability and creep mechanisms. Reports of room-temperature grain growth—a common feature unique to highly pure NC metals—has been reported numerous times and is in stark contrast to the growth that takes place at much higher temperatures in coarse-grained metals (such as in the experiments on NC Cu reported in ref. 32). Considerable research has been undertaken to address this specific limitation, culminating in two main methods, one based on thermodynamics and the other on kinetics^{33,34}. The thermodynamic approach deals with reducing the excess free energy of the grain boundaries through solute segregation, whereas the kinetic approach deals with reducing grain-boundary mobility. The thermodynamic approach is considered to be more promising because it attenuates the driving force for grain growth, whereas kinetic approaches based on solute drag, chemical ordering and Zener pinning continually fight against the system reaching equilibrium with an Arrhenius temperature dependence.

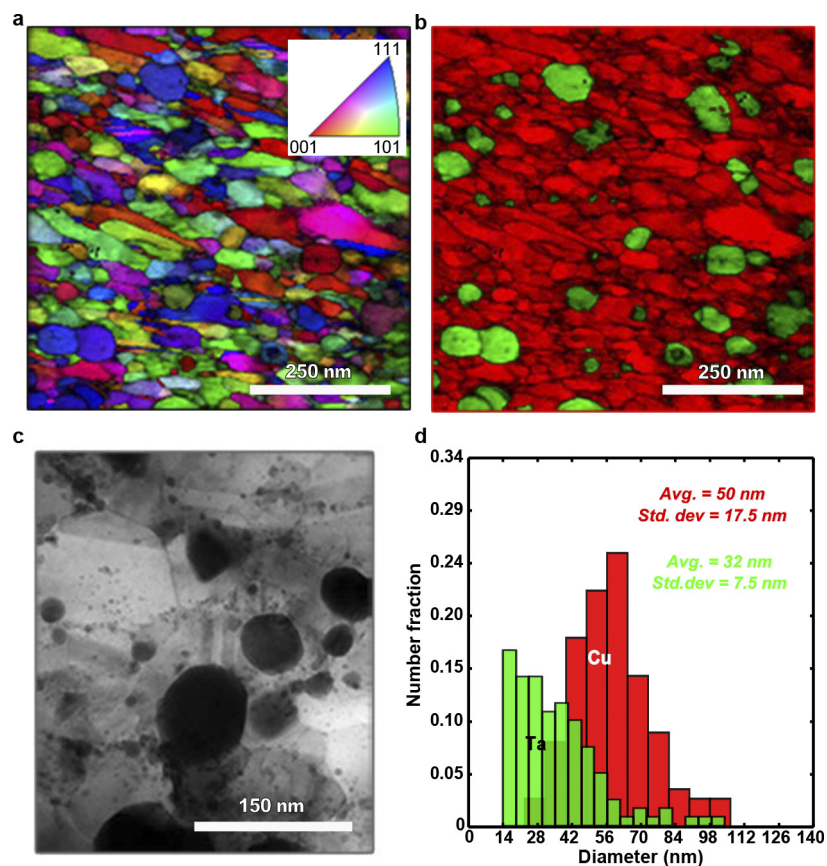
In light of these two competing mechanisms there has been much discussion on which method may provide a more successful path in bringing about the realization of commercially available bulk NC metals¹⁰. In many cases, such debate has been fostered by the fact that it is not always possible to fully separate or delineate the contributions of these two competing stabilization mechanisms in preventing

grain growth in NC metals. For instance, thermodynamic stabilization of NC grain size involves examining the energetic penalty associated with the high volume fraction of the grain boundaries, and the possibility of solute segregation driving this associated excess free energy to zero³⁵. However, intertwined in this scenario are the kinetic aspects of solute drag, and its role in reducing grain growth in this thermodynamic stabilization construct, which has been an area of active research³⁴. Additionally, the precipitation of secondary, solute-rich phases have been experimentally observed to disrupt the stabilization set in place by the thermodynamic mechanism¹⁰. However, recent research has shown that the occurrence of phase separation or precipitation does not necessarily mean that a stabilized NC system does not exist¹⁰. Recent theoretical work³⁶ predicts the existence of stable duplex systems, wherein both grain-boundary segregation and phase separation occurs, resulting in a stable NC grain size (that is, grain-boundary energy of zero) and a precipitate structure coexisting with one another. These types of microstructures are currently under investigation. In reference to these particular immiscible NC Cu–10 at% Ta alloys, the nature of their thermal decomposition and formation of an extremely high density of clusters, occurring primarily along grain boundaries, gives rise to an unusually stable microstructure. Additionally, we have reason to believe the exact mechanisms of Zener pinning in this system may be more complicated than conventional theory¹⁸. Nevertheless, the NC Cu–10 at% Ta alloys primarily stabilized kinetically by small-scale coherent clusters shown here provide a design route to the development of advanced structural materials for various applications including high-strength, high-temperature applications. Many of the processing and consolidation challenges that have haunted NC metals are now more fully understood, opening the door for bulk NC metals and parts to be produced. This has been made possible by the advancement of thermodynamic, kinetic and thermo-kinetic methods of stabilizing their microstructures. The Cu–Ta family of alloys are currently one of the few systems that have been shown to retain NC grain sizes in a fully dense part, allowing the study of these microstructures under extreme environments. To the best of our knowledge, our work represents the first time a stable NC metal has been studied under conditions of high-temperature, high-stress creep.

21. Segal, V. M. Materials processing by simple shear. *Mater. Sci. Eng. A* **197**, 157–164 (1995).
22. Furukawa, M., Horita, Z., Nemoto, M. & Langdon, T. G. Review: processing of metals by equal-channel angular pressing. *J. Mater. Sci.* **36**, 2835–2843 (2001).
23. Zhu, Y. T. & Lowe, T. C. Observations and issues on mechanisms of grain refinement during ECAP process. *Mater. Sci. Eng. A* **291**, 46–53 (2000).
24. Vincent, R. & Midgley, P. A. Double conical beam-rocking system for measurement of integrated electron diffraction intensities. *Ultramicroscopy* **53**, 271–282 (1994).
25. Eshelby, J. D. The determination of the elastic field of an ellipsoidal inclusion, and related problems. *Proc. R. Soc. Lond. A* **241**, 376–396 (1957).
26. Nabarro, F. R. N. Dislocations in a simple cubic lattice. *Proc. Phys. Soc.* **59**, 256–272 (1947).
27. Herring, C. Diffusional viscosity of a polycrystalline solid. *J. Appl. Phys.* **21**, 437–445 (1950).
28. Lagneborg, R. & Bergman, B. The stress/creep rate behaviour of precipitation-hardened alloys. *Met. Sci.* **10**, 20–28 (1976).
29. Du, Q., Faber, V. & Gunzburger, M. Centroidal Voronoi tessellations: applications and algorithms. *SIAM Rev.* **41**, 637–676 (1999).
30. Bhatia, M. A., Mathaudhu, S. N. & Solanki, K. N. Atomic-scale investigation of creep behavior in nanocrystalline Mg and Mg–Y alloys. *Acta Mater.* **99**, 382–391 (2015).
31. Yamakov, V., Wolf, D., Phillpot, S. & Gleiter, H. Grain-boundary diffusion creep in nanocrystalline palladium by molecular-dynamics simulation. *Acta Mater.* **50**, 61–73 (2002).
32. Chokshi, A. H., Rosen, A., Karch, J. & Gleiter, H. On the validity of the Hall-Petch relationship in nanocrystalline materials. *Scr. Metall.* **23**, 1679–1683 (1989).
33. Schiøtz, J., Di Tolla, F. D. & Jacobsen, K. W. Softening of nanocrystalline metals at very small grain sizes. *Nature* **391**, 561–563 (1998).
34. Koch, C. C. Structural nanocrystalline materials: an overview. *J. Mater. Sci.* **42**, 1403–1414 (2007).
35. Chookajorn, T., Murdoch, H. A. & Schuh, C. A. Design of stable nanocrystalline alloys. *Science* **337**, 951–954 (2012).
36. Murdoch, H. A. & Schuh, C. A. Estimation of grain boundary segregation enthalpy and its role in stable nanocrystalline alloy design. *J. Mater. Res.* **28**, 2154–2163 (2013).

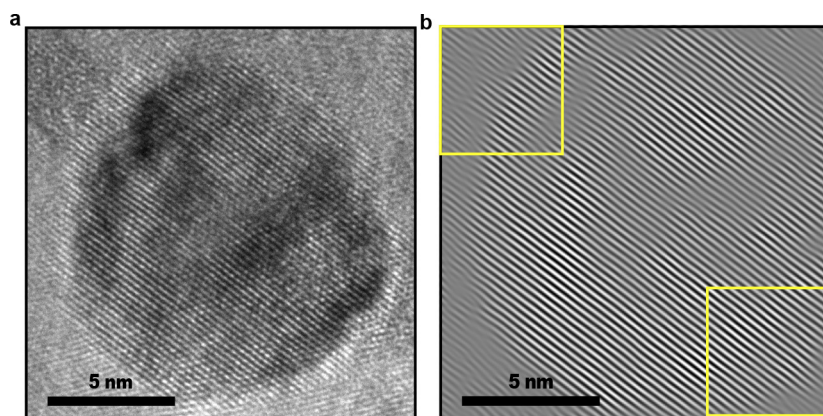


Extended Data Figure 1 | X-ray diffraction analysis. X-ray diffraction plot showing Cu and Ta reflections from the as-received NC Cu-10 at% Ta sample processed at 700 °C.



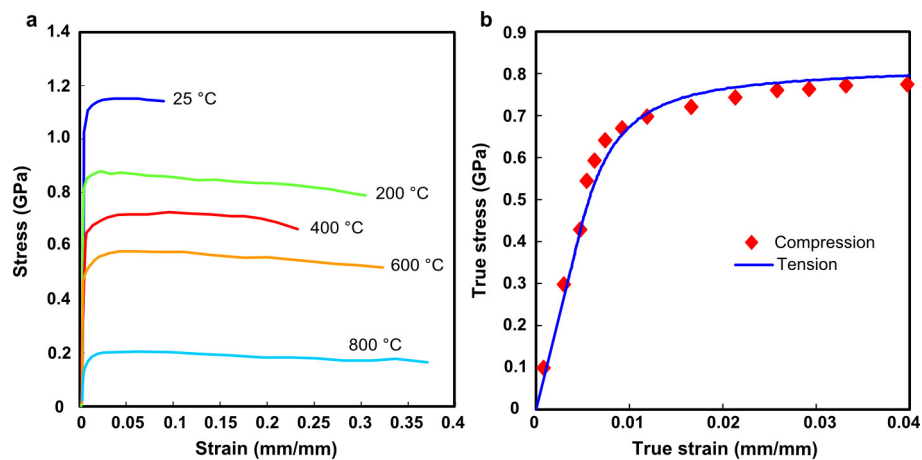
Extended Data Figure 2 | As-received microstructure characterization. TEM characterization of as-received NC Cu-10 at% Ta. **a**, Precession diffraction TEM micrograph revealing orientation detail of nanometre-sized grains. **b**, Phase map of the region in the TEM micrograph in **a** showing two

distinct phases: Cu (red) and Ta (green). **c**, Bright-field TEM micrograph showing the microstructure. **d**, Size distributions of Cu and Ta grains determined from 300 grains.



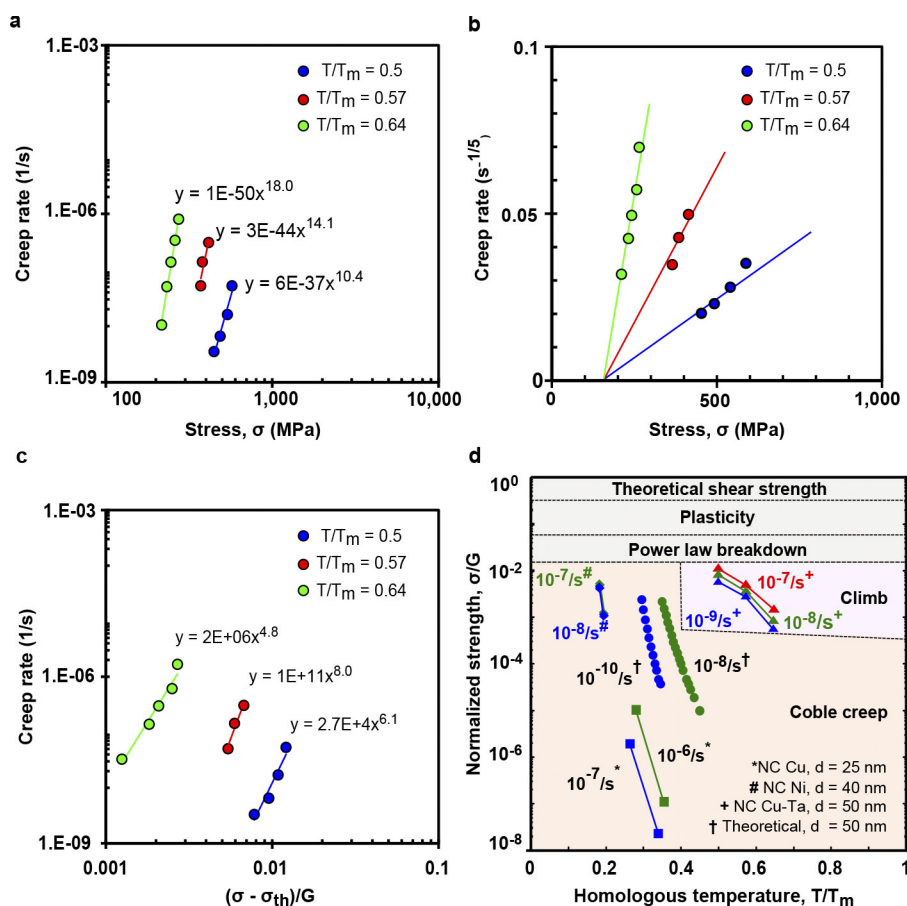
Extended Data Figure 3 | TEM characterization of a Ta-rich nanocluster. **a**, HR-TEM micrograph of a Ta-rich nanocluster with misfit lattice dislocations. **b**, Inverse fast Fourier transform image of the

nanocluster in **a** highlighting the misfit dislocations (indicated by the yellow boxes) across the interface indicative of the presence of misfit strain.



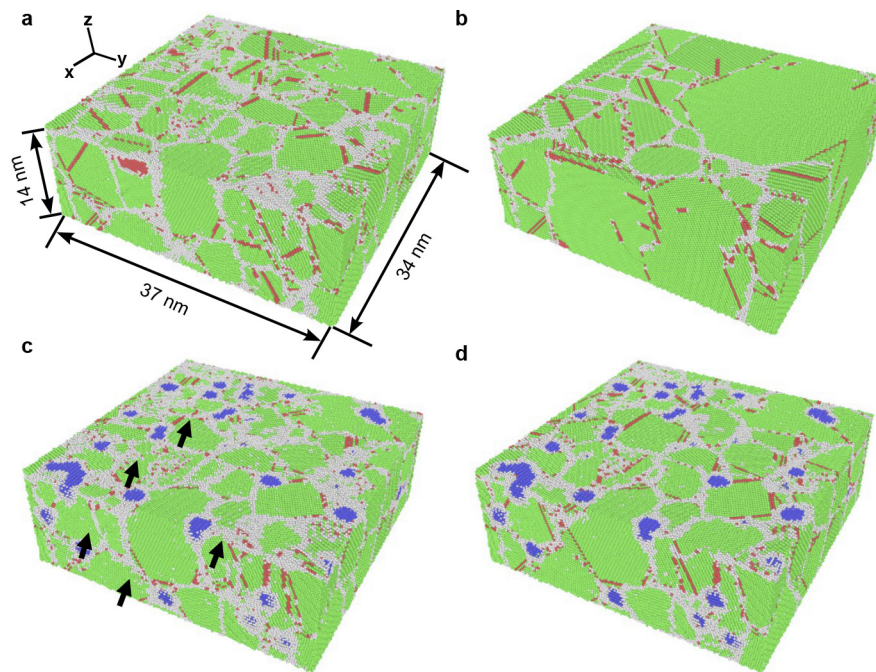
Extended Data Figure 4 | Mechanical behaviour of NC Cu-10 at% Ta at quasi-static strain rates. Stress-strain response of NC Cu-10 at% Ta samples. **a**, Compressive stress-strain curve tested at a strain rate of $8 \times 10^{-4} \text{ s}^{-1}$ at various temperatures⁵. **b**, Tension-compression curve

for samples tested at a strain rate of $1 \times 10^{-3} \text{ s}^{-1}$. The blue line and red diamonds corresponds to tensile and compressive data, respectively. The curves indicate behaviour ranging from elastic to nearly plastic with no strain hardening; tension-compression asymmetry is absent.



Extended Data Figure 5 | Creep response of NC Cu–10 at% Ta with and without threshold correction. **a**, Creep rate versus applied stress without threshold correction for various temperatures. **b**, Creep rate versus applied stress with a stress exponent of five for various temperatures, used to extract the threshold stress. **c**, Creep rate versus normalized stress, obtained by subtracting the threshold stress (σ_{th}) from the applied stress (σ) for various temperatures. The red, green and blue circles in **a–c** correspond to NC Cu–10 at% Ta tested at 400 °C, 500 °C and 600 °C,

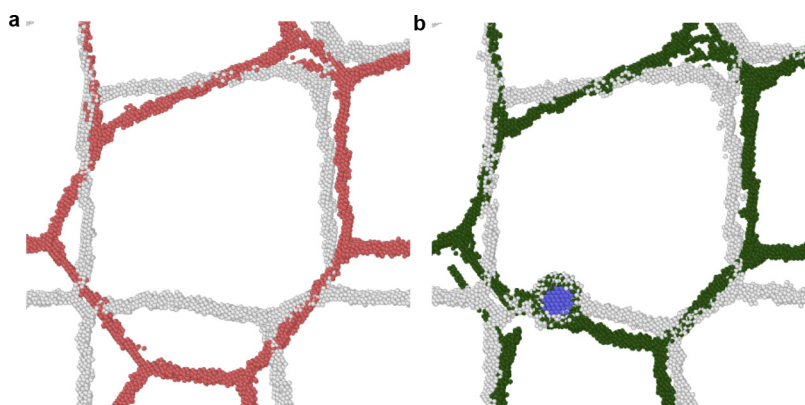
respectively. **d**, Theoretical deformation map with threshold-corrected data ($n = 5$) of the NC Cu–10 at% Ta alloy, along with experimental creep rates of NC Cu and Ni (ref. 2). Theoretical Coble creep rates for a grain size $d = 50$ nm (circles), experimental creep rates² in nanocrystalline copper (squares; $d = 25$ nm) and nickel (diamonds; $d = 40$ nm), and the creep rate we found for NC Cu–10 at% Ta (triangles; $d = 50$ nm) are plotted.



Extended Data Figure 6 | Atomistic models used for creep study.

a, Initial microstructure of pure NC Cu. **b**, Microstructure of pure NC Cu after 5 ns. **c**, Initial microstructure of NC Cu-10 at% Ta. **d**, Microstructure of NC Cu-10 at% Ta after 5 ns. In all panels, green atoms are grain-interior

(face-centred cubic) atoms, red atoms are stacking faults (hexagonal close-packed) atoms, white atoms are grain-boundary (other) atoms and blue atoms are Ta (body-centred cubic) particles.



Extended Data Figure 7 | Isolated single grain illustrating the effect of Ta nanoclusters on grain-boundary motion. a, b, 2D atomistic slices of a single grain obtained from pure NC Cu (a) and NC Cu–10 at% Ta (b) highlighting the effect of nanoclusters in coarsening. White atoms

represent the initial grain-boundary configurations and blue atoms represent Ta; red and green atoms represent the extent of coarsening associated with the plastic deformation under constant load and temperature conditions.

Enhanced electrocatalytic CO₂ reduction via field-induced reagent concentration

Min Liu^{1*}, Yuanjie Pang^{2*}, Bo Zhang^{1,3*}, Phil De Luna^{4*}, Oleksandr Voznyy¹, Jixian Xu¹, Xueli Zheng^{1,5}, Cao Thang Dinh¹, Fengjia Fan¹, Changhong Cao², F. Pelayo García de Arquer¹, Tina Saberi Safaei¹, Adam Mepham⁶, Anna Klinkova⁷, Eugenia Kumacheva⁷, Tobin Filleter², David Sinton², Shana O. Kelley^{6,8,9} & Edward H. Sargent¹

Electrochemical reduction of carbon dioxide (CO₂) to carbon monoxide (CO) is the first step in the synthesis of more complex carbon-based fuels and feedstocks using renewable electricity^{1–7}. Unfortunately, the reaction suffers from slow kinetics^{7,8} owing to the low local concentration of CO₂ surrounding typical CO₂ reduction reaction catalysts. Alkali metal cations are known to overcome this limitation through non-covalent interactions with adsorbed reagent species^{9,10}, but the effect is restricted by the solubility of relevant salts. Large applied electrode potentials can also enhance CO₂ adsorption¹¹, but this comes at the cost of increased hydrogen (H₂) evolution. Here we report that nanostructured electrodes produce, at low applied overpotentials, local high electric fields that concentrate electrolyte cations, which in turn leads to a high local concentration of CO₂ close to the active CO₂ reduction reaction surface. Simulations reveal tenfold higher electric fields associated with metallic nanometre-sized tips compared to quasi-planar electrode regions, and measurements using gold nanoneedles confirm a field-induced reagent concentration that enables the CO₂ reduction reaction to proceed with a geometric current density for CO of 22 milliamperes per square centimetre at –0.35 volts (overpotential of 0.24 volts).

This performance surpasses by an order of magnitude the performance of the best gold nanorods, nanoparticles and oxide-derived noble metal catalysts. Similarly designed palladium nanoneedle electrocatalysts produce formate with a Faradaic efficiency of more than 90 per cent and an unprecedented geometric current density for formate of 10 milliamperes per square centimetre at –0.2 volts, demonstrating the wider applicability of the field-induced reagent concentration concept.

The Gibbs free energy (ΔG) diagrams obtained from density functional theory (DFT) calculations on gold (Au) surface models of various facets at 298 K, 1 atm and 0 V versus reversible hydrogen electrode (RHE) are given in Fig. 1 (see also Extended Data Fig. 1 and Extended Data Table 1a–c), showing that adsorbed K⁺ ions lower the thermodynamic energy barrier for reaction for all facets. On the Au(111) gold surface, the adsorbed K⁺ stabilizes the COOH* and CO* intermediates by 0.89 eV and 0.24 eV, respectively (Fig. 1a). On Au(100) and Au(110), it stabilizes the rate-determining COOH* intermediate⁸ by 0.66 eV and 0.69 eV, respectively (Fig. 1b, c). On the under-coordinated Au(211) facet, K⁺ similarly stabilizes COOH* and CO* (Fig. 1d). We further note that in the presence of adsorbed K⁺, a greater electron density is found on the carbon of the COOH*

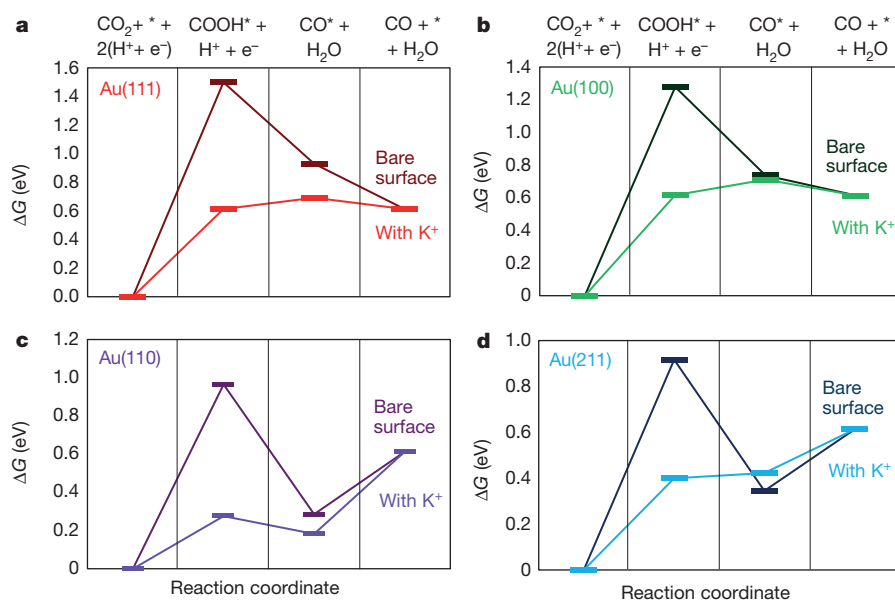


Figure 1 | Thermodynamic barriers for the CO₂-to-CO reduction reaction on Au surface under conditions with and without K⁺. Gibbs free energy ΔG diagrams of the electrochemical reduction of CO₂ to CO on Au(111) (a), Au(100) (b), Au(110) (c) and Au(211) (d) facets in the presence of adsorbed K⁺ and in the absence of adsorbed K⁺.

¹Department of Electrical and Computer Engineering, University of Toronto, 35 St George Street, Toronto, Ontario M5S 1A4, Canada. ²Department of Mechanical and Industrial Engineering, University of Toronto, 5 King's College Road, Toronto, Ontario M5S 3G8, Canada. ³Department of Physics, East China University of Science and Technology, 130 Meilong Road, Shanghai 200237, China. ⁴Department of Materials Science and Engineering, University of Toronto, 184 College Street, Toronto, Ontario M5S 3E4, Canada. ⁵Tianjin Key Laboratory of Composite and Functional Materials, School of Materials Science and Engineering, Tianjin University, Tianjin 300072, China. ⁶Institute of Biomaterials and Biomedical Engineering, University of Toronto, 164 College Street, Toronto, Ontario M5S 3G9, Canada. ⁷Department of Chemistry, University of Toronto, 80 St George Street, Toronto, Ontario M5S 3H6, Canada. ⁸Department of Pharmaceutical Sciences, Leslie Dan Faculty of Pharmacy, University of Toronto, 144 College Street, Toronto, Ontario M5S 3M2, Canada. ⁹Department of Biochemistry, University of Toronto, 1 King's College Circle, Toronto, Ontario M5S 1A8, Canada.

*These authors contributed equally to this work.

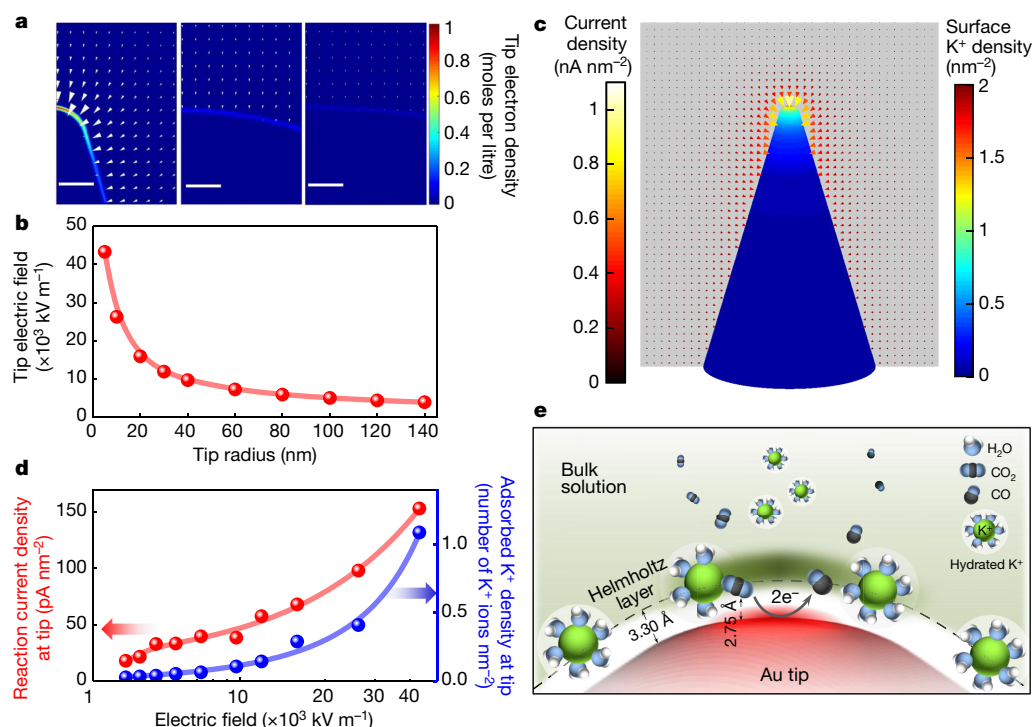


Figure 2 | Computed electric field, K^+ concentration and current density near the tip of an electrode as a function of tip radius. **a**, Free electron density distribution on the surface of electrodes is shown as a colour map. Electrostatic field distribution around the electrode is shown as a group of arrows, where the size and direction of each arrow represent the magnitude and direction of the field at the arrow's spatial position. The tip radius of the structure in each panel is 5 nm (left), 60 nm (middle) and 140 nm (right). Scale bars represent 5 nm. **b**, Electrostatic field intensity at the electrode tip increases as the tip radius decreases. **c**, Surface K^+ density and current density distributions on the surface of Au needles. The tip radius is 5 nm. **d**, Adsorbed K^+ and reaction current density as functions of the electric field intensity at the tip. **e**, A schematic showing how K^+ ions on the gold surface help CO_2 molecule adsorption.

intermediate, suggesting a stronger C–Au bond (Extended Data Fig. 1e) and further indicating that adsorbed cations modulate the CO_2 reduction reaction.

Ab initio molecular dynamics simulations reveal that the presence of K^+ reduces the mean square displacement of CO_2 relative to the surface by a factor of 2.3, compared to the system with no K^+ (Extended Data Fig. 1f), converging to approximately 2.5 \AA^2 regardless of facet (Extended Data Fig. 1g). The nearest-neighbour C–Au distances obtained from simulated radial distribution function peaks are 2.75 \AA and 3.25 \AA in the presence and absence of K^+ , respectively (Extended Data Fig. 1h and Extended Data Table 1d). Further, the interaction energy of CO_2 on the Au surface with K^+ is consistently smaller (Extended Data Fig. 1i).

These results suggest that locally concentrating cations at reactive sites could enhance CO_2 electroreduction. As high-curvature structures are known to concentrate electric fields that can affect ion concentrations, we used a finite-element numerical method to explore the prospects of tip-enhanced nanometre-scale field intensification and cation concentration. Cones with rounded tips were used to represent sharp electrode tips immersed in an electrolyte, with their tip-concentrated electron density (Fig. 2a) increasing as the electrodes sharpen. The locally enhanced electrostatic field is generated by, and points to, the locally concentrated free electron density on the surface of the electrodes (arrows in Fig. 2a). It originates from the migration of free electrons to the regions of the sharpest curvature on a charged metallic electrode, a consequence of electrostatic repulsion¹². Tip sharpening from a radius of 140 nm to 5 nm enhances electrostatic field intensity (Fig. 2b) at the tip of the electrode, at the CO_2/CO equilibrium potential (-0.11 V), by one order of magnitude.

To estimate the quantitative impact of the electric field on the surface-adsorbed cation concentration, we used a Gouy–Chapman–Stern model (Extended Data Fig. 2a and Methods) to map the surface-adsorbed K^+ ion density in the Helmholtz layer of the electrical double layer directly adjacent to the electrode surface (Fig. 2c). This indicates a 20-fold increased surface-adsorbed K^+ ion concentration at the Au needle tip due to locally enhanced electrostatic field (Fig. 2d), while a sixfold increase in the bulk K^+ concentration in the electrolyte only doubles the field-induced K^+ ion concentration near the electrode

(Extended Data Fig. 2b). Furthermore, increasing the applied cathode potential tenfold, from -0.11 V to -1.1 V (where the CO_2 reduction reaction— CO_2RR —is no longer selective because it competes against H_2 evolution) only doubles the field-induced K^+ ion concentration (Extended Data Fig. 2c). With concentrated K^+ , CO_2 quickly (in 0.5 ps) stabilizes on the Au sharp features (Extended Data Fig. 2d) and CO_2RR mostly occurs at the Au tips (Fig. 2c and Extended Data Fig. 2e), with the effect projected to increase the reduction current by two orders of magnitude (Fig. 2d). These results, taken together, point to field-induced reagent concentration (FIRC) as a means of enhancing CO_2RR appreciably (Fig. 2e).

To probe experimentally the predictions, we used electrodeposition as a convenient and scalable means of preparing desired electrodes¹³ with a suite of tip radii (Fig. 3 and Extended Data Fig. 3a) ranging from large-diameter particles (radius of curvature of about 140 nm) to intermediate-diameter rods (radius of curvature of about 60 nm) to high-curvature nanoneedles (radius of curvature of about 5 nm). Electrochemical roughness factors were measured via two electrochemical methods^{9,14}, providing the values of 52, 33 and 12 for Au needle, rod and particle electrodes, respectively (Extended Data Fig. 3b, c and Extended Data Table 2). X-ray diffraction confirms that all micro- and nano-structures comprise a regular (uncompressed) gold lattice (Extended Data Fig. 3d). X-ray photoelectron spectroscopy and O K-edge X-ray absorption spectra show features characteristic of Au^0 and none attributable to oxide (Extended Data Figs 3d and 4a, b). High-resolution transmission electron microscopy (TEM) and the corresponding local electron energy loss spectroscopy (EELS, Extended Data Fig. 4c, d) show no Au adatoms or local Au oxide on the tips of the Au needles.

Kelvin probe atomic force microscopy confirmed that electric fields are highest for the needles and lowest for large particles (Fig. 3c, g, k). Secondary Au nanoparticle electrodeposition preferentially occurs at the tip of Au needles (Fig. 3d), decreases on Au rods and almost disappears on Au particles (Extended Data Fig. 5f). Au needles have the largest electric-field-induced locally adsorbed K^+ concentration under performance-testing conditions (Fig. 3h), with conductive atomic force microscopy proving that the nanoscale local current at Au needle tips is higher than the current on Au rods and particles (Fig. 3l and Extended Data Fig. 2f). These results all support the

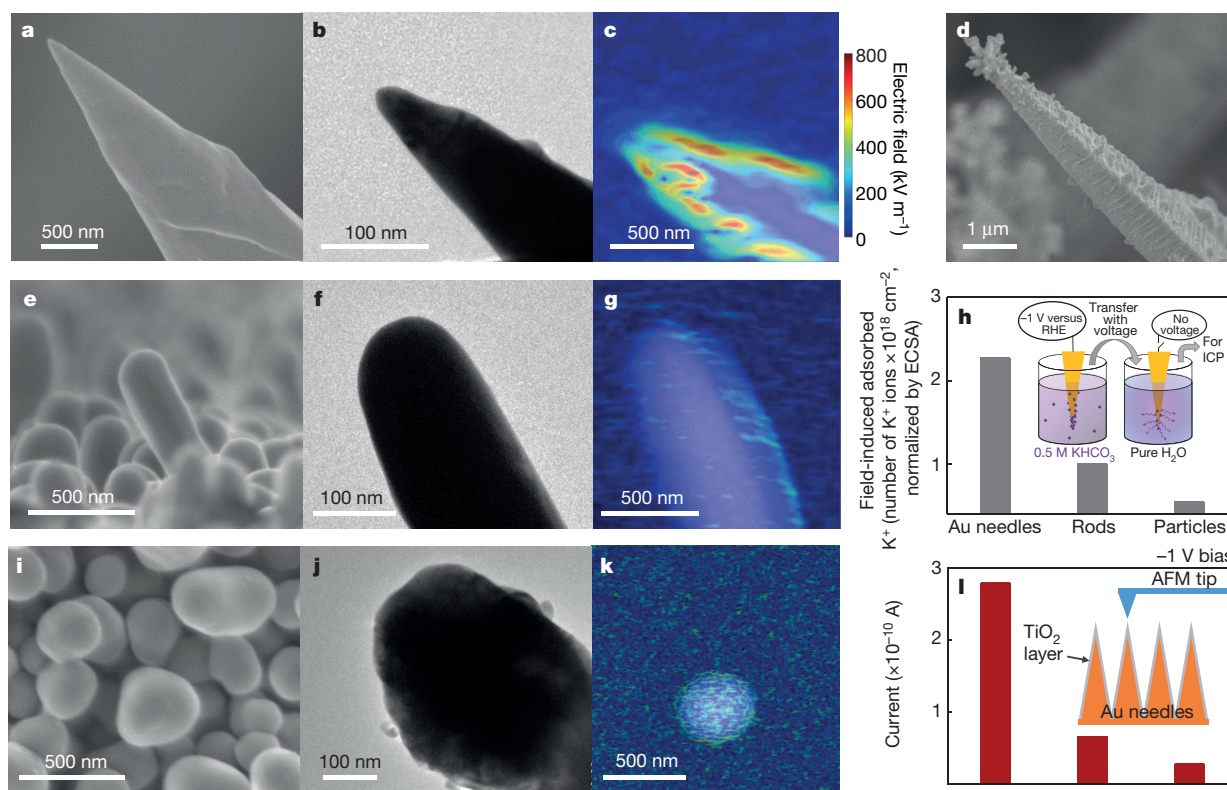


Figure 3 | Physical characterization of Au tips, rods and particles. **a, e, i,** Scanning electron microscopy (SEM) images; **b, f, j,** TEM images; **c, g, k,** Electric field distribution of Au needles, rods and particles deduced using Kelvin probe atomic force microscopy. **d,** SEM image of Au needle with secondarily deposited Au particles. **h,** ECSA-normalized field-induced concentration of adsorbed K^+ on Au needles, rods and particles.

The concentration of K^+ was measured via inductively coupled plasma (ICP) optical emission spectrometry. The inset shows the process of measuring the field-induced adsorbed K^+ . **i,** Current on a single Au needle, rod and particle with a thin TiO_2 insulator layer at a bias of -1 V . The inset shows the current measurement conditions.

local presence of large electric fields and the FIRC effect at the Au needle tips.

To validate the predicted enhancement of CO_2RR by FIRC, we explored the CO_2 reduction activity of Au needles, rods and particles in CO_2 -saturated 0.5 M KHCO_3 ($\text{pH } 7.2$). Products were quantified using gas chromatography. The linear sweep voltammetry curves exhibit a clear reduction peak for the Au needles in the range -0.30 V to -0.50 V (Fig. 4a), whereas Au rods and particles only give smooth current–voltage curves. Notably, Au needles exhibited a stable total geometric current density (j_{tot}) of approximately 15 mA cm^{-2} at a potential of -0.35 V (corresponding to an overpotential η_{CO} of 0.24 V for CO production^{4,9}) during 8 h of continuous reaction (Fig. 4b). The Faradaic efficiency for CO production was nearly quantitative ($>95\%$) throughout the electrocatalytic process. No obvious changes in the morphology, crystal structure and surface state were observed after long-term CO_2RR (Extended Data Fig. 5a), indicating that the Au needles are stable under electrocatalytic conditions. Au rods and particles exhibited j_{tot} values of approximately 0.7 mA cm^{-2} and 0.1 mA cm^{-2} after 8 h of reaction. Their Faradaic efficiencies for CO were about 25% and 3% , respectively. The approximately 20-fold difference in CO_2RR current between Au needles and Au rods agrees with the increase in surface-adsorbed K^+ ion concentration and current density predicted by theory (Fig. 2d).

The differences in CO_2 reduction activity among Au needles, rods and particles were more pronounced at lower overpotentials. At -0.3 V ($\eta_{\text{CO}} = 0.19 \text{ V}$), Au needles exhibited $j_{\text{tot}} \approx 7 \text{ mA cm}^{-2}$ over the course of 8 h of electrolysis and about 90% Faradaic efficiency for CO production (Extended Data Fig. 3h), while Au particles exhibited very low current densities ($<0.05 \text{ mA cm}^{-2}$) and exclusively H_2 evolution. Au rods also showed a low current density of about 0.1 mA cm^{-2} and a

very poor $\sim 3\%$ selectivity for CO formation (Extended Data Fig. 3h). No detectable CO_2 reduction was observed for Au rods at applied potentials closer to RHE than -0.3 V , whereas Au needles continued to reduce CO_2 —at -0.2 V ($\eta_{\text{CO}} = 0.09 \text{ V}$), Au needles gave a j_{tot} value of about 0.6 mA cm^{-2} with a Faradaic efficiency of about 40% for CO (Extended Data Fig. 3i), and at the exceptionally low potential of -0.18 V ($\eta_{\text{CO}} = 0.07 \text{ V}$) the CO product remained readily detectable using gas chromatography (Faradaic efficiency of about 6%). A summary of CO_2 reduction Faradaic efficiencies at potentials between -0.18 V and -0.5 V for the different systems is given in Fig. 4c.

Intrinsic performances can be compared by considering the geometric and the electrochemical active surface area (ECSA)-normalized partial current densities for CO production (geometric current density j_{CO}) versus applied potential for the three classes of electrodes (Extended Data Fig. 3j, k). Once current is renormalized by the ECSA, the j_{CO} value measured at -0.35 V on Au needles is 63 times higher than on rods and 112 times higher than on particles (Extended Data Fig. 3j, k), indicating higher intrinsic CO_2RR activities for Au needles.

Tafel analysis (Fig. 4d) gives for Au needles, rods and particles a slope of 42 mV dec^{-1} , 80 mV dec^{-1} and 96 mV dec^{-1} , respectively. Previous studies suggest that during two-electron CO_2RR , the first one-electron step of CO_2 to COOH^* or $\text{CO}_2^{\cdot-}$ intermediates is determining the rate for the combined process^{6–8} and hence the Tafel slope. The Tafel slope measured for the gold particles of 96 mV dec^{-1} agrees well with prior reports^{9,15} (114 mV dec^{-1} and 129 mV dec^{-1}), whereas the much lower Tafel slope of 42 mV dec^{-1} obtained for the needles indicates a faster first-electron transfer step^{9,16,17} and confirms the superiority of Au needles in CO_2 reduction. These observations agree with the FIRC picture of Au needles concentrating CO_2 at the electrode and with modelled Tafel slopes that assume cathodic charge transfer coefficients are 0.95 ,

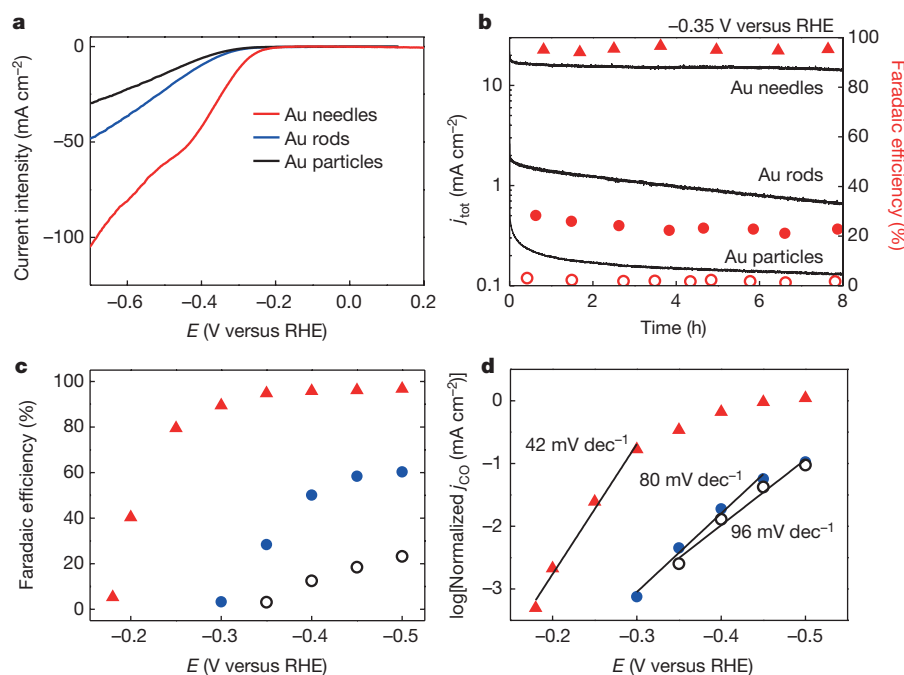


Figure 4 | **CO₂ reduction performances on Au needles, rods and particles in 0.5 M KHCO₃, pH 7.2.** **a**, Current–voltage curves on Au needles, rods and particles obtained from the linear sweep voltammetry scans. Scan rate, 10 mV s^{−1}. **b**, CO₂ reduction activity of Au needles, rods and particles at −0.35 V versus RHE. Total current density (left axis) versus time and CO Faradaic efficiency (right axis) versus time. **c**, CO Faradaic efficiencies on Au needles, rods and particles at different applied potentials. **d**, ECSA-normalized CO production partial current density versus potential on Au needles, rods and particles.

0.49 and 0.43 for Au needles, rods and particles, respectively (Extended Data Fig. 2f).

To assess the energy barriers, we studied the effect of temperature on the performance of the catalysts (Extended Data Fig. 3l–n) and found the rate constants to follow the Arrhenius relationship. The electrochemical activation energies of 72 kJ mol^{−1}, 44 kJ mol^{−1} and 21 kJ mol^{−1} extracted for particles, rods and needles from the slope of their Arrhenius plots at $\eta = 240$ mV (see the insets of Extended Data Fig. 3l–n) are comparable to those reported previously (Extended Data Table 2)^{18–22}, with the lowest value for Au needles highlighting the dominant role of thermodynamics in the CO₂RR.

To probe charge transfer processes occurring at electrode/solution interfaces, we obtained electrochemical impedance (Z) spectra (Extended Data Fig. 3g and Extended Data Table 2). For Au needles, the semicircle diameter of the Nyquist plot is much smaller, reflecting an acceleration of the charge transfer process. Thus, both charge separation and the kinetics of charge transfer on Au needles are improved.

Specific facets, grain boundaries, metastable surfaces, corner and edge sites have all previously been invoked as structural features that enhance CO₂-to-CO electroreduction activity^{8,9,23–26}. To test whether these could account for the higher intrinsic activity of the Au needles, in one experiment we overcoated the needles with additional Au thin layers; in a second we annealed the Au needles in vacuum; in a third, we used etching solutions designed to expose (111) facets preferentially²⁷; and in a fourth, we used plasma bombardment to produce fresh reconstructions (Extended Data Fig. 5b–e). None of these surface-manipulating treatments affected the activity of Au needles (all performance remained within 10% of the original), so we conclude that the primary benefits of the needles in CO₂RR are not related to the details of surface faceting nor of atomic-scale structure.

We systematically varied the nanostructure morphologies. When we grew dendritic Au nanoleaves, we obtained a low j_{CO} value and Faradaic efficiency, which we attribute to their high radius (50–500 nm) of curvature (Extended Data Fig. 6). When we dulled the needles by electrodepositing a thick additional Au layer, the large-radius (50–100 nm) nanoparticles covering the tips led to notably worsened performance in CO₂RR even though the ECSA had been increased 1.7-fold owing to the attached nanoparticles (Extended Data Fig. 5g and h). Electrochemical oxidation of Au needles further dulls the tips of the needles and the CO₂RR performance was further decreased even

though oxide-derived Au was generated on the surface (Extended Data Fig. 4e–j)⁹.

Further experiments performed at different K⁺ concentrations confirm that CO₂RR performance increases with K⁺ concentration (Extended Data Fig. 7a, b). Notably, Au needles exhibited a j_{CO} of approximately 22 mA cm^{−2} at −0.35 V after 8 h by using a saturated KHCO₃ solution (Extended Data Fig. 7c). The long-term steady-state CO₂ reduction current density for the highest-performing Au needle morphology is over one order of magnitude higher at −0.35 V than for any previously reported CO₂ reduction catalysts in aqueous solution with inorganic electrolyte (Extended Data Table 3); this comparison takes into account the best nanostructured oxide-derived gold electrodes^{9,23–25}.

We also carried out CO₂RR experiments in solutions without alkali cations, such as CO₂ saturated NH₄HCO₃ solution and H₂O (Extended Data Fig. 7e, f). The results show that the current density and selectivity decreased substantially, and only H₂ was generated in pure H₂O and was accompanied by a very small current density. The results, taken together, confirm that the FIRC dictates the CO₂RR rate.

To demonstrate the universality of the FIRC, we prepared palladium (Pd) needles and tested their CO₂RR performance (Extended Data Fig. 8). The obtained Pd needles exhibited enhanced CO₂-to-formate conversion compared with that of rods and particles. Pd needles exhibited a stable geometric current density j_{formate} of approximately 10 mA cm^{−2} at −0.2 V over the course of 20 h in 0.5 M KHCO₃ solution (Extended Data Fig. 8). Faradaic efficiency for formate generation was nearly quantitative (>91%) throughout the electrocatalytic process. This formate production current density on Pd needles is over three times higher at −0.2 V than the previously reported CO₂-to-formate catalysts in aqueous solution (Extended Data Table 3)^{2,28}, confirming that the FIRC concept can be extended to other CO₂RR systems.

The sharp-tip enhancement effect may have contributed to previous studies identifying particularly active CO₂RR sites at corners⁸ and ridges^{24,25}, since such sites are locally high-curvature regions. It remains to be explored whether it will be effective in industrial electrolyzers operating at current densities of 300 mA cm^{−2} (that is, with reaction rates ten times faster than studied here), but enhanced control over the density of sharp tips and use of high bulk CO₂ concentrations could enhance CO₂RR rates further towards the goal of industrial electrosynthesis of carbon-based fuels. In a wider electrochemistry context, the tip-enhanced field phenomenon can

be extended to concentrate the reagents locally in other reactions and as such suggests a general principle for the design of efficient electrodes for catalysis.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 November 2015; accepted 15 June 2016.

Published online 3 August 2016.

- Costentin, C., Drouet, S., Robert, M. & Saveant, J. M. A local proton source enhances CO₂ electroreduction to CO by a molecular Fe catalyst. *Science* **338**, 90–94 (2012).
- Gao, S. *et al.* Partially oxidized atomic cobalt layers for carbon dioxide electroreduction to liquid fuel. *Nature* **529**, 68–71 (2016).
- Lin, S. *et al.* Covalent organic frameworks comprising cobalt porphyrins for catalytic CO₂ reduction in water. *Science* **349**, 1208–1213 (2015).
- Rosen, B. A. *et al.* Ionic liquid-mediated selective conversion of CO₂ to CO at low overpotentials. *Science* **334**, 643–644 (2011).
- Qiao, J. L., Liu, Y. Y., Hong, F. & Zhang, J. J. A review of catalysts for the electroreduction of carbon dioxide to produce low-carbon fuels. *Chem. Soc. Rev.* **43**, 631–675 (2014).
- Hori, Y. in *Modern Aspects of Electrochemistry* Vol. 42 (eds Vayenas, C. *et al.*) 89–189 (Springer, 2008).
- Lu, Q. *et al.* A selective and efficient electrocatalyst for carbon dioxide reduction. *Nat. Commun.* **5**, 3242 (2014).
- Back, S., Yeom, M. S. & Jung, Y. Active sites of Au and Ag nanoparticle catalysts for CO₂ electroreduction to CO. *ACS Catal.* **5**, 5089–5096 (2015).
- Chen, Y., Li, C. W. & Kanan, M. W. Aqueous CO₂ reduction at very low overpotential on oxide-derived Au nanoparticles. *J. Am. Chem. Soc.* **134**, 19969–19972 (2012).
- Varela, A. S., Kroschel, M., Reier, T. & Strasser, P. Controlling the selectivity of CO₂ electroreduction on copper: the effect of the electrolyte concentration and the importance of the local pH. *Catal. Today* **260**, 8–13 (2016).
- Kokoszka, B., Jarrah, N. K., Liu, C., Moore, D. T. & Landskron, K. Supercapacitive swing adsorption of carbon dioxide. *Angew. Chem. Int. Ed.* **53**, 3698–3701 (2014).
- Jackson, J. D. *Classical Electrodynamics* Vol. 3, 75–79 (Wiley, 1962).
- Soleymani, L., Fang, Z., Sargent, E. H. & Kelley, S. O. Programming the detection limits of biosensors through controlled nanostructuring. *Nat. Nanotechnol.* **4**, 844–848 (2009).
- Das, J. & Kelley, S. O. Tuning the bacterial detection sensitivity of nanostructured microelectrodes. *Anal. Chem.* **85**, 7333–7338 (2013).
- Noda, H., Ikeda, S., Yamamoto, A., Einaga, H. & Ito, K. Kinetics of electrochemical reduction of carbon-dioxide on a gold electrode in phosphate buffer solutions. *Bull. Chem. Soc. Jpn.* **68**, 1889–1895 (1995).
- Gileadi, E. *Electrode Kinetics for Chemists, Engineers, and Materials Scientists* Ch. 1 (VCH, 1993).
- Chen, Y. H. & Kanan, M. W. Tin oxide dependence of the CO₂ reduction efficiency on tin electrodes and enhanced activity for tin/tin oxide thin-film catalysts. *J. Am. Chem. Soc.* **134**, 1986–1989 (2012).
- Morris, A. J., McGibbon, R. T. & Bocarsly, A. B. Electrocatalytic carbon dioxide activation: the rate-determining step of pyridinium-catalyzed CO₂ reduction. *ChemSusChem* **4**, 191–196 (2011).
- Peebles, D. E., Goodman, D. W. & White, J. M. Methanation of carbon dioxide on nickel(100) and the effects of surface modifiers. *J. Phys. Chem.* **87**, 4378–4387 (1983).
- Wang, W., Wang, S., Ma, X. & Gong, J. Recent advances in catalytic hydrogenation of carbon dioxide. *Chem. Soc. Rev.* **40**, 3703–3727 (2011).
- Ghuman, K. K. *et al.* Illuminating CO₂ reduction on frustrated Lewis pair surfaces: investigating the role of surface hydroxides and oxygen vacancies on nanocrystalline In₂O₃-x(OH)_y. *Phys. Chem. Chem. Phys.* **17**, 14623–14635 (2015).
- Schmeier, T. J., Dobereiner, G. E., Crabtree, R. H. & Hazari, N. Secondary coordination sphere interactions facilitate the insertion step in an iridium(III) CO₂ reduction catalyst. *J. Am. Chem. Soc.* **133**, 9274–9277 (2011).
- Feng, X. F., Jiang, K. L., Fan, S. S. & Kanan, M. W. Grain-boundary-dependent CO₂ electroreduction activity. *J. Am. Chem. Soc.* **137**, 4606–4609 (2015).
- Zhu, W. *et al.* Monodisperse Au nanoparticles for selective electrocatalytic reduction of CO₂ to CO. *J. Am. Chem. Soc.* **135**, 16833–16836 (2013).
- Zhu, W. L. *et al.* Active and selective conversion of CO₂ to CO on ultrathin Au nanowires. *J. Am. Chem. Soc.* **136**, 16132–16135 (2014).
- Zhang, Q. & Wang, H. Facet-dependent catalytic activities of Au nanoparticles enclosed by high-index facets. *ACS Catal.* **4**, 4027–4033 (2014).
- Mettela, G. & Kulkarni, G. U. Facet selective etching of Au microcrystallites. *Nano Res.* **8**, 2925–2934 (2015).
- Min, X. & Kanan, M. W. Pd-catalyzed electrohydrogenation of carbon dioxide to formate: high mass activity at low overpotential and identification of the deactivation pathway. *J. Am. Chem. Soc.* **137**, 4701–4708 (2015).

Acknowledgements This work was supported by the Ontario Research Fund: Research Excellence programme, the Natural Sciences and Engineering Research Council (NSERC) of Canada, the CIFAR Bio-Inspired Solar Energy programme and a University of Toronto Connaught grant. B.Z. acknowledges funding from Shanghai Municipal Natural Science Foundation (14ZR1410200) and the National Natural Science Foundation of China (21503079). We thank X. Lan, P. Kanjanaboos, G. Walters, L. Levina, R. Wolowiec, D. Kopilovic, E. Palmiano, T. Burdyny and J. Tam from the University of Toronto for Au electron beam deposition, liquid products testing, AFM testing, TEM EELS measurements, discussions and additional aids during the course of study, Y. Tian from the King Abdullah University of Science and Technology for electrode preparation assistance, and M. Bajdich, L. D. Chen and K. Chan from Stanford University for advice on DFT calculations. This work has also benefited from the Spherical Grating Monochromator beamlines at the Canadian Light Source. DFT calculations were performed on the IBM BlueGene Q supercomputer with support from the Southern Ontario Smart Computing Innovation Platform (SOSCIP).

Author Contributions E.H.S., S.O.K. and D.S. supervised the project. M.L., Y.P. and B.Z. designed and carried out all the experiments and COMSOL simulations. P.D.L. and O.V. carried out the DFT simulation. All authors discussed the results and assisted during manuscript preparation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.H.S. (ted.sargent@utoronto.ca).

METHODS

DFT calculations. DFT calculations were performed on $3 \times 3 \times 3$ slabs of Au(111), Au(110), Au(100), and Au(211) using the generalized gradient approximation exchange correlation functional of ref. 29. All DFT simulations were performed with the Vienna *ab initio* Simulation Package (VASP)³⁰ using the projector augmented wave method³¹. The projector augmented wave pseudopotentials^{31,32} were used to calculate the interaction between ions and electrons in a plane wave basis set with a cut-off energy of 500 eV and a $5 \times 5 \times 1$ Monkhorst–Pack mesh³³ used for k -point sampling and a Fermi-level smearing of 0.1 eV. Spin polarization was included as it has been previously shown to be important for binding energies on gold nanoparticles and surfaces³⁴. The surface slabs were modelled with 10 Å of vacuum and dipole corrections were implemented. Structural optimizations were performed with the Broyden–Fletcher–Goldfarb–Shanno⁴⁵ (BFGS) algorithm until the maximum force was less than 0.02 eV per atom, with the surface slab fully relaxed. Once the slab models were optimized, all subsequent thermodynamic calculations were performed with the bottom two layers fixed.

All thermodynamic properties were calculated using the open-source atomic simulation environment suite of programs³⁵. The Gibbs free energies were calculated at 298 K and 1 atm as outlined below:

$$G = H - T\Delta S = E_{\text{DFT}} + E_{\text{ZPE}} + \int_0^{298} C_v dT - T\Delta S$$

where E_{DFT} is the DFT-optimized total energy, E_{ZPE} is the zero-point vibrational energy, $\int_0^{298} C_v dT$ is the heat capacity, T is the temperature, and ΔS is the entropy.

Gas-phase molecules such as CO_2 and H_2 were treated using the ideal gas approximation, whereas adsorbates were treated using a harmonic approximation. The DFT-calculated energy for CO_2 was corrected by 0.45 eV, a common adjustment to account for an overestimation by DFT³⁶. The change in Gibbs free energy ΔG between reaction steps of the CO_2 to CO reaction coordinate was calculated from the computational hydrogen electrode model³⁷. Additionally, the binding energy was calculated from DFT-optimized structures as follows: $E_{\text{binding}} = E_{\text{CO}_2^*} - (E_{\text{Au}} + E_{\text{CO}_2})$ where $E_{\text{CO}_2^*}$ is the energy of the system with CO_2 proximate to the Au surface, E_{Au} is the energy of the gold surface (with and without K^+ for the respective cases), and E_{CO_2} is the gas-phase energy of CO_2 .

Charge density analysis was performed from the electron density as calculated from DFT. The volume slice was visualized in Visual Molecular Dynamics (VMD, <http://www.ks.uiuc.edu/Research/vmd/>) with an isovalue of 0.5 (ref. 38). Bader partial atomic charges were calculated using the Bader Charge Analysis code as maintained by the Henkelman group³⁹.

Ab initio molecular dynamics simulations. All *ab initio* molecular dynamics simulations on $6 \times 6 \times 5$ slabs of Au(111), Au(110), Au(100) and Au(211) were performed within the DFT framework as mentioned above with a cut-off energy of 400 eV and gamma k -point sampling of the Brillouin zone. The electronic self-consistent loop was considered to be converged if the energy difference was lower than 10^{-5} eV, at which point the molecular dynamics would continue to the next time step. A canonical ensemble using a Nosé–Hoover thermostat was used with a constant temperature of 300 K. Fermi-smearing was used owing to the presence of the Au(111) metal surface, with 0.2 eV used as the width of smearing. A 5-ps total simulation run was performed with 1-ps equilibration and 4-ps production runs and a time step of 1 fs for 5,000 steps. An ensemble average of the radial distribution function and mean square displacement was obtained from 25 unique runs starting from the same initial configuration in order to better sample the binding event of CO_2 to Au.

COMSOL Multiphysics simulations. Free electron density on the electrodes, as well as the electric field and potassium ion density within the vicinity of the electrodes was simulated using the COMSOL Multiphysics finite-element-based solver (<https://www.comsol.com/>). The ‘Electric currents’ module was used to solve the free electron density on the electrode under a specific electrode bias potential. Electric field \mathbf{E} was computed as the opposite gradient of the electric potential V as follows: $\mathbf{E} = -\nabla V$.

The electric conductivity of the electrode (gold) was taken to be $4.42 \times 10^7 \text{ S m}^{-1}$ (ref. 40). The electrolyte conductivity was assumed to be 10 S m^{-1} . Charge density ρ was computed using Gauss’s law for electric field: $\rho = \varepsilon_r \varepsilon_0 \nabla \cdot \mathbf{E}$, where ε_0 represents the dielectric function for a vacuum, and ε_r represents the dielectric function of the materials, and equals 78 for the electrolyte and 1 for gold.

In this work, the electrical double layer was modelled using the Gouy–Chapman–Stern model, which consists of a Helmholtz layer and a diffusion layer (illustrated in Extended Data Fig. 2a). The Helmholtz layer consists of a monolayer of surface-adsorbed hydrated cation on the electrode surface, which speeds up the CO_2 RR. The diffusion layer consists of both cations and anions, which freely

diffuse in the electrolyte and form concentration gradients towards and away from the electrode surface. The diffusion layer was established as the result of a dynamic equilibrium between electrostatic forces and diffusion (that is, the ‘entropic forces’). The ‘Electrostatics’ and the ‘Transport of diluted species’ modules were combined to solve the potassium ion density in the electrical double layer. The Poisson–Nernst–Planck equations were solved in the steady state:

$$\nabla^2 V = \begin{cases} 0 & d < d_H \\ (c_K - c_{\text{HCO}_3})F & d > d_H \end{cases}$$

$$\nabla \cdot \left(D \nabla c_i + \frac{Dz_i e}{k_B T} c_i \nabla V \right) = 0$$

Here d is the distance from the electrode surface into the electrolyte, and d_H is the thickness of the Helmholtz layer, which is taken as the radius of a hydrated potassium ion (0.33 nm)⁴¹. That is, $d < d_H$ within the Helmholtz layer, and $d > d_H$ in the diffusion layer. c_i with $i \in \{\text{K}^+, \text{HCO}_3^-\}$ are the concentrations of the potassium or bicarbonate ion, z_i are the valencies of both ions, e is the elementary charge, k_B is Boltzmann constant, the absolute temperature T was taken as 297.3 K. The diffusion coefficients D of the potassium ion, the bicarbonate ion, and the proton in water were taken to be $2.14 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$, $7.02 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$ and $7.10 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$ (ref. 42). Two-dimensional axisymmetric models were built to represent the three-dimensional nanoneedle, nanorod and nanoparticle structures used in this work. Triangular meshes were used for all simulations. Meshes were set to be the densest at the surface of the electrodes, where the element size was 0.17 nm. In other parts of the model where less precision is required, for example, in the bulk electrolyte, the maximum element size was 20 nm.

The electrochemical module in COMSOL was used to obtain the CO_2 to CO reaction current density using the Butler–Volmer equation:

$$i = i_0 \left[\exp \left(\frac{\alpha_a n F \eta}{RT} \right) - \exp \left(-\frac{\alpha_c n F \eta}{RT} \right) \right]$$

where α_a and α_c are the dimensionless anodic and cathodic charge transfer coefficients, respectively, $n = 2$ is the number of electrons involved in the electrode reaction, F is the Faraday constant, R is the universal gas constant, and T is temperature, taken to be 293.15 K. The exchange current density i_0 obeys the Arrhenius law:

$$i_0 \propto \exp \left(-\frac{E_a}{k_B T} \right)$$

where k_B is the Boltzmann constant, and E_a is the activation energy of the CO_2 to CO reaction, which was experimentally obtained to be 0.59 eV without K^+ , and 0.21 eV with K^+ .

Preparation of gold needle, rod, particle and leaf electrodes. Gold electrodes were prepared through an electrodeposition process using a solution containing HAuCl_4 (99.99% Sigma) and HCl (TraceSELECT) solution¹³. The concentration of HCl was fixed at 0.5 mol l^{-1} (M). Gold-coated slides (for characterization, EMF Corporation) and carbon paper (for CO_2 RR performance measurement, Toray TGP-H-060, purchased from Fuel Cell Store) were used as substrates ($0.1\text{--}0.3 \text{ cm}^2$). The Au needle electrode was formed using a 160 mM HAuCl_4 solution and direct current potential amperometry at -400 mV for 300 s. Au particle, rod and leaf electrodes were formed using direct current potential amperometry at -250 mV with 13 mM, 26 mM and 40 mM HAuCl_4 solutions for 1,200 s, 900 s and 600 s, respectively.

Preparation of palladium needle, rod and particle electrodes. Pd needles were synthesized by a two-step potential square wave electrodeposition in a solution of 2 mM K_2PdCl_6 in 0.5 M H_2SO_4 (ref. 43) on an Autolab PGSTAT302N potentiostat. In the first step, E_1 , T_1 , E_2 and T_2 were 0.8 V, 0.05 s, -0.7 V and 0.02 s, respectively. The number of square waves was 1,200. For the second step E_1 , T_1 , E_2 and T_2 were 0.6 V, 0.005 s, 0.25 V and 0.005 s, respectively. The number of square waves was 100,000. For the preparation of Pd rods, E_1 was set at 0.2 V with the number of square waves being 50,000 in the second step. All other parameters remained the same as for the Pd needles. For preparation of Pd particles, only the first step for Pd needle deposition was applied and the square wave number was set to 50,000.

Au secondary electrodeposition. After washing with deionized water and drying, Au needles, rods and particles were coated with a thin layer of Au by electrodeposition in a solution of 20 mM HAuCl_4 and 0.5 M HClO_4 . The secondary deposition was performed by using direct current potential amperometry at -400 mV for 30 s.

Surface, grain boundaries and Au oxide investigation. To exclude the influence of the surface states, such as surface facets, corner sites and edge sites, a uniform Au

thin layer with thickness of 10 nm was deposited on Au needles, rods and particles by using electron-beam deposition with a rate of 0.4 Å s^{-1} . The surface of Au needles were also etched by using CuCl_2 solution (5 mM)²⁷. Briefly, Au nanoneedles was immersed in a vial containing 15 ml of CuCl_2 solution (5 mM). The vial was then heated to 70°C using an oil bath and kept at that temperature for 1 h. The etched Au nanoneedles obtained were washed with a copious amount of water and dried at room temperature.

To investigate the influence of grain boundaries and metastable surface states, the Au needle electrode was annealed at 140°C in vacuum for 24 h and treated with plasma bombardment for 1 h (50 W, argon atmosphere). To investigate the influence of Au oxide, Au needles were oxidized in aqueous 0.5 M H_2SO_4 at 1.5 V versus Ag/AgCl for 10 h.

ECSA measurement. We used two methods to estimate the ECSA of Au needles, rods and particles. In the first we integrated the reduction peak area obtained from a cyclic voltammogram in 50 mM H_2SO_4 (ref. 44). In the second, we measured the charge associated with the stripping of an underpotential-deposited Cu monolayer⁹. In the first method, cyclic voltammograms from 0 V to 1.5 V (versus Ag/AgCl) at a scan rate of 50 mV s^{-1} were acquired repeatedly until the traces converged. In the forward scan, a monolayer of chemisorbed oxygen is formed and then it is reduced in the reverse scan. The surface area was calculated by integrating the reduction peak (0.9 V versus Ag/AgCl) to obtain the reduction charge. The reduction charge per microscopic unit area has been experimentally determined to be $448 \mu\text{C cm}^{-2}$.

In the underpotential-deposited method, the electrode was immersed in a 0.50 M H_2SO_4 solution containing 100 mM CuSO_4 continuously purged with N_2 . Cyclic voltammograms from 0.83 V to 0.483 V (versus Ag/AgCl) at a scan rate of 50 mV s^{-1} were acquired repeatedly until traces converged. The anodic stripping waves at 0.403 V versus Ag/AgCl were integrated. The factor used to convert the stripping charge to surface area was $92.4 \mu\text{C cm}^{-2}$. The error of the results obtained from these two methods are within 5%, indicating an accurate estimation of ECSA.

Electric-field-induced adsorbed K^+ . Electric-field-induced adsorbed K^+ was performed in 0.5 M KHCO_3 solution. Au needles, rods and particles were run in the solution at -1 V . Once the running time reached 120 s, the electrode was directly raised above the solution. After removing the applied potential, the electrodes were immersed in 10 ml pure water and any adsorbed K^+ on the Au needles was released into the pure water. Then, the amount of K^+ in the water was checked using an inductively coupled plasma optical emission spectrometer (ICP-OES, Agilent Dual-View 720 with a charge-coupled device (CCD) detector for full wavelength coverage between 167 nm and 785 nm). The obtained results were normalized by using ECSA.

Characterization. The structural characteristics of the prepared samples were measured by powder X-ray diffraction at room temperature on a MiniFlex600 instrument with a copper target ($\lambda = 1.54056 \text{ Å}$). The morphologies of the prepared Au electrodes were investigated using SEM on a Hitachi SU-8230 apparatus and TEM on a Hitachi HF-3300 instrument with an acceleration voltage of 200 kV. Compositions were studied by X-ray photoelectron spectroscopy (model 5600, Perkin-Elmer). The binding energy data were calibrated with reference to the C 1s signal at 284.5 eV. Kelvin probe atomic force microscopy images were obtained using an Asylum Research MFP-3D instrument. Electrostatic field E around the electrodes was calculated to have the opposite gradient of the electric potential raw data V from Kelvin probe atomic force microscopy imaging: $E = -\nabla V$. Currents on single Au needle, rod and particle were measured by using a Cypher ES instrument with a conductive model. Before current measurement, a 10-nm-thick layer of TiO_2 was deposited on the surface of Au needles, rods and particles using a Picosun R200 atomic layer deposition system. Soft X-ray absorption measurements were performed at the Spherical Grating Monochromator beamline of the Canadian Light Source in Saskatoon.

Electrocatalytic reduction of CO_2 . All CO_2 reduction experiments were performed using a three-electrode system connected to an electrochemical workstation (Autolab PGSTAT302N). Ag/AgCl (with saturated KCl as the filling solution) and platinum mesh were used as reference and counter electrodes, respectively.

Electrode potentials were converted to the reversible hydrogen electrode (RHE) reference scale using $E_{\text{RHE}} = E_{\text{Ag/AgCl}} + 0.197 \text{ V} + 0.0591 \times \text{pH}$.

The electrolyte was 0.5 M KHCO_3 saturated with CO_2 with pH of 7.2. The experiments were performed in a gas-tight two-compartment H-cell separated by an ion exchange membrane (Nafion117). The electrolyte in the cathodic compartment was stirred at a rate of 300 r.p.m. during electrolysis. CO_2 gas was delivered into the cathodic compartment at a rate of 5.00 standard cubic centimeters per minute (s.c.c.m.) and was routed into a gas chromatograph (PerkinElmer Clarus 600). The gas chromatograph was equipped with a Molecular Sieve 5A capillary column and a packed Carboxen-1000 column. Argon (Linde, 99.999%) was used as the carrier gas. The gas chromatograph columns led directly to a thermal conductivity detector to quantify hydrogen and a flame ionization detector equipped with a methanizer to quantify carbon monoxide. The partial current densities of CO and H_2 production were calculated from the gas chromatograph peak areas as below⁹:

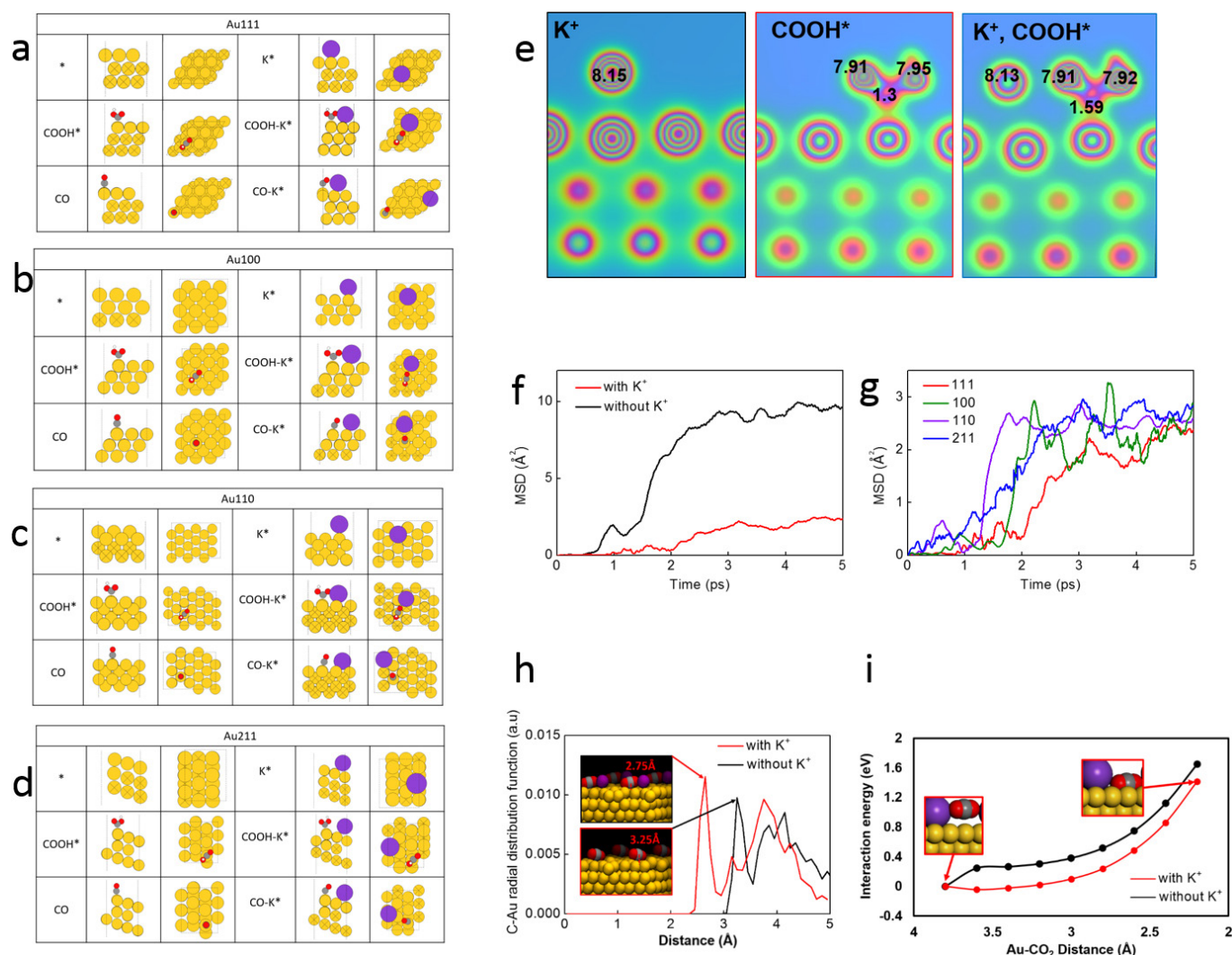
$$j_{\text{CO}} = \frac{\text{peak area}}{\alpha} \times \text{flow rate} \times \frac{2Fp_0}{RT} \times (\text{electrode area})^{-1}$$

$$j_{\text{H}_2} = \frac{\text{peak area}}{\beta} \times \text{flow rate} \times \frac{2Fp_0}{RT} \times (\text{electrode area})^{-1}$$

where α and β are conversion factors for CO and H_2 respectively based on calibration of the gas chromatograph with standard samples, $p_0 = 1.013 \text{ bar}$ and $T = 300 \text{ K}$.

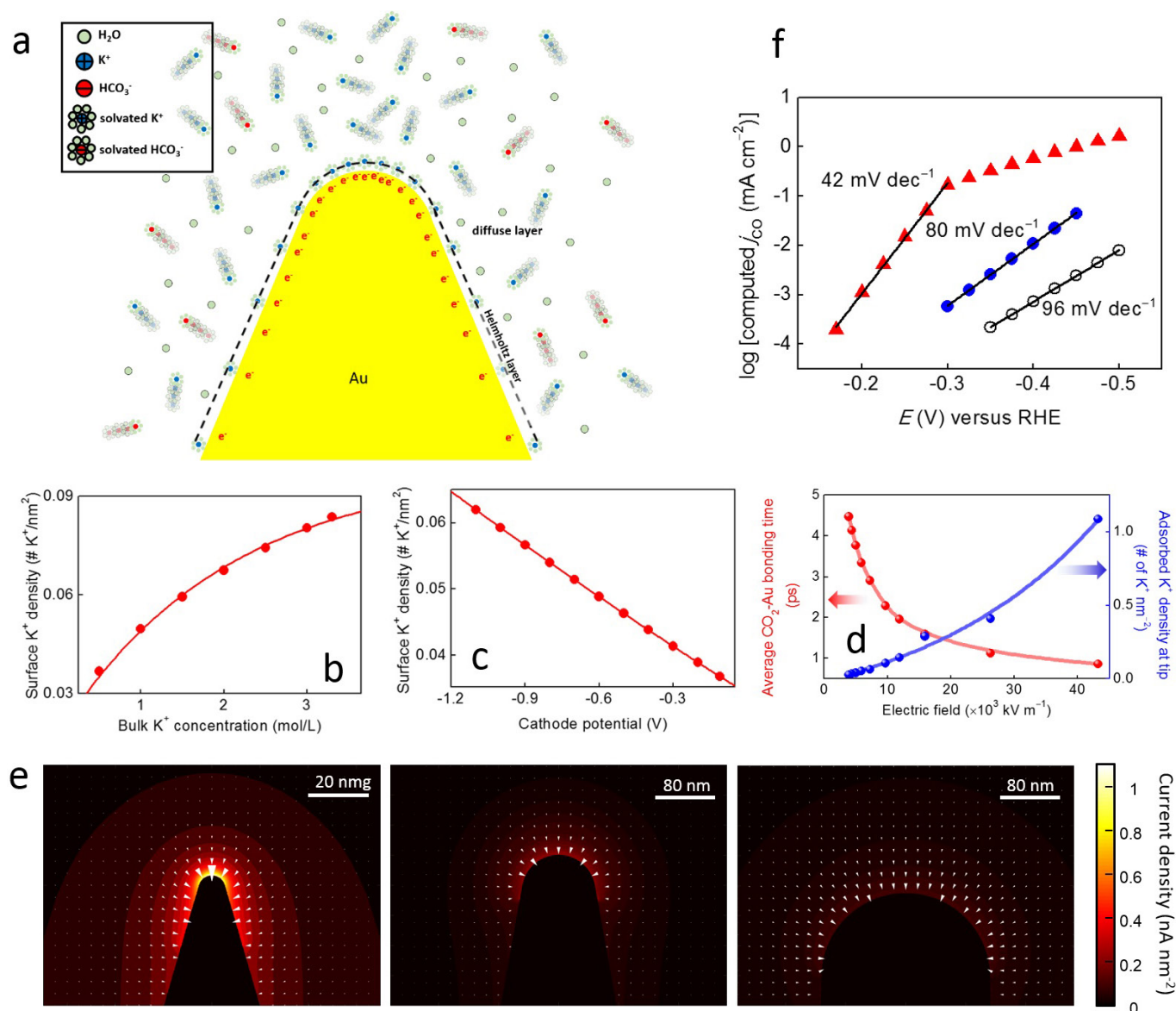
Formate was quantified on a gas chromatograph with mass spectrometry (PerkinElmer Clarus 600 GC-MS System). Assuming that two electrons are needed to produce one formate molecule, the Faradaic efficiency was calculated as $2Fn_{\text{formate}}/Q = 2Fn_{\text{formate}}/(It)$, where F is the Faraday constant, I is the current, t is the running time and n_{formate} is the total amount of produced formate (in moles).

29. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
30. Kresse, G. & Furthmüller, J. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
31. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
32. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
33. Monkhorst, H. J. & Pack, J. D. Special points for Brillouin-zone integrations. *Phys. Rev. B* **13**, 5188–5192 (1976).
34. Kleis, J. *et al.* Finite size effects in chemical bonding: from small clusters to solids. *Catal. Lett.* **141**, 1067–1071 (2011).
35. Bahn, S. R. & Jacobsen, K. W. An object-oriented scripting interface to a legacy electronic structure code. *Comput. Sci. Eng.* **4**, 56–66 (2002).
36. Studt, F. *et al.* The mechanism of CO and CO_2 hydrogenation to methanol over Cu-based catalysts. *ChemCatChem* **7**, 1105–1111 (2015).
37. Peterson, A. A., Abild-Pedersen, F., Studt, F., Rossmeisl, J. & Nørskov, J. K. How copper catalyzes the electroreduction of carbon dioxide into hydrocarbon fuels. *Energy Environ. Sci.* **3**, 1311–1315 (2010).
38. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
39. Tang, W., Sanville, E. & Henkelman, G. A grid-based Bader analysis algorithm without lattice bias. *J. Phys. Condens. Matter* **21**, 084204 (2009).
40. Serway, R. A. & Jewett, J. W. *Principles of Physics* Vol. 1, 602 (Saunders College Pub., 1998).
41. Wang, H. & Pilon, L. Accurate simulations of electric double layer capacitance of ultramicroelectrodes. *J. Phys. Chem. C* **115**, 16711–16719 (2011).
42. Friedman, A. M. & Kennedy, J. W. The self-diffusion coefficients of potassium, cesium, iodide and chloride ions in aqueous solutions. *J. Am. Chem. Soc.* **77**, 4499–4501 (1955).
43. Meng, H., Wang, C., Shen, P. K. & Wu, G. Palladium thorn clusters as catalysts for electrooxidation of formic acid. *Energy Environ. Sci.* **4**, 1522–1526 (2011).
44. Bin, X. M., Sargent, E. H. & Kelley, S. O. Nanostructuring of sensors determines the efficiency of biomolecular capture. *Anal. Chem.* **82**, 5928–5931 (2010).
45. Head, J. D. & Zerner, M. C. A Broyden–Fletcher–Goldfarb–Shanno optimization procedure for molecular geometries. *Chem. Phys. Lett.* **122**, 264–270 (1985).



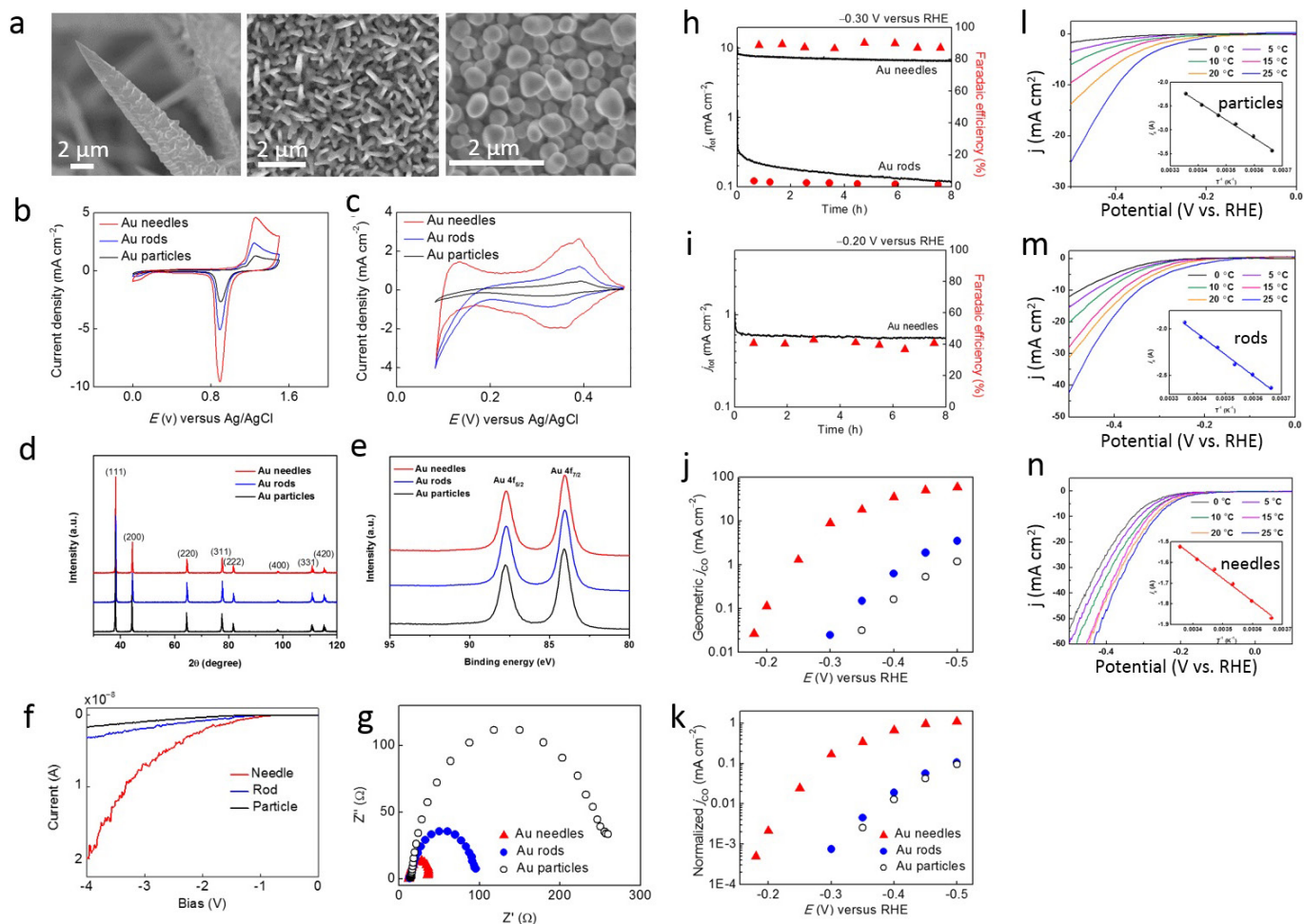
Extended Data Figure 1 | Optimized structure for Au facets and data calculated with or without K⁺. **a–d**, Optimized structures. **a**, Au(111) facet. **b**, Au(100) facet. **c**, Au(110) facet. **d**, Au(211) facet. Included are the optimized positions of the adsorbates COOH and CO without and with the presence of an adsorbed K⁺ (purple). **e**, Volume slice of calculated charge densities. Bader partial atomic charges are indicated in black with and without K⁺. In the presence of K⁺ the Bader partial atomic charge on the carbon of COOH* has increased from 1.3 to 1.59 suggesting higher electron density and thus a stronger C–Au bond. **f**, Calculated average mean square displacement of CO₂ on Au(111) surface with and without K⁺ in the system. This ensemble average shows CO₂ is more diffuse

without a K⁺ cation to facilitate CO₂ surface binding. **g**, Mean square displacement of CO₂ on Au(111), Au(110), Au(100) and Au(211) surface in the presence of K⁺. It was found that regardless of facet the mean square displacement of CO₂ converges to about 2.5 Å². **h**, Calculated C–Au radial distribution function of CO₂ to Au(111) from an ensemble average of 25 *ab initio* molecular dynamics simulations (5 ps) shows CO₂ is closer to the surface of gold on average in the presence of K⁺ than without K⁺. **i**, Calculated interaction energy of CO₂ vary with C–Au distance under the conditions with or without K⁺. The interaction energy is consistently less in the presence of an adsorbed K⁺ (red) than without K⁺ (black).



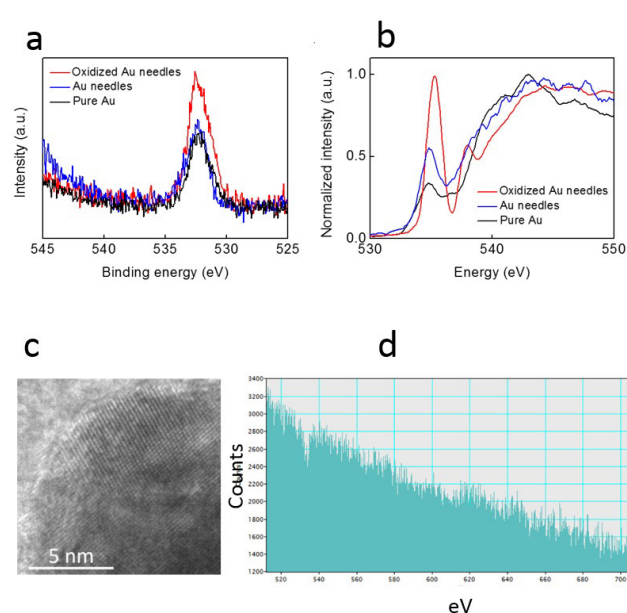
Extended Data Figure 2 | Electrochemical simulation model and results. **a**, Schematic of the Gouy-Chapman-Stern electrical double layer model. **b**, Field-induced surface K^+ ion concentration as a function of bulk K^+ ion concentration. **c**, Field-induced surface K^+ ion concentration as a function of electrode potential (versus RHE). **d**, Required CO_2 -Au bonding time versus electric field. With concentrated K^+ , CO_2 quickly (in 0.5 ps) stabilizes on the Au sharp features and remains there for the remainder of the simulation run. **e**, Current density distributions on

the surface of Au structures. The tip radius is 5 nm. The tip radius of the structure in each panel is: 5 nm (top), 60 nm (middle) and 140 nm (bottom). Arrows are magnified 2 \times in the middle panel and 4 \times in the bottom panel for the purpose of clarity. **f**, Simulated Tafel plots for needles (tip radius 5 nm), rods (tip radius 60 nm), particles (tip radius 140 nm). Simulated data was fitted to the experimental data with fitting parameter cathodic charge transfer coefficient being 0.95 (needles), 0.49 (rods), and 0.43 (particles).

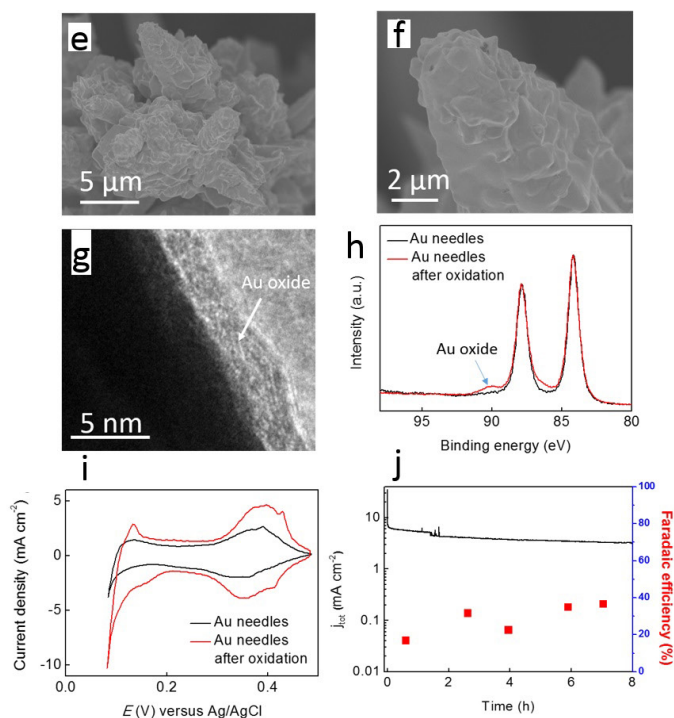


Extended Data Figure 3 | Additional physical characterization, CO_2 reduction and kinetic analyses of Au samples. **a**, Morphologies for Au tips (left), rods (middle) and particles (right) imaged by SEM. **b**, **c**, ECSA measurement. **b**, Cyclic voltammograms in 50 mM H_2SO_4 . Scan rate 50 mV s^{-1} . **c**, Underpotential Cu deposition and anodic stripping waves. The electrolyte solution was 100 mM CuSO_4 in 0.50 M H_2SO_4 . Scan rate 50 mV s^{-1} . **d**, X-ray diffraction patterns for all of the electrodes exhibited peaks at the expected positions for an ideal Au lattice, indicating no uniform expansion or compression of the unit cell. **e**, X-ray photoelectron spectroscopy exhibited the expected peaks for Au^0 but no peaks attributable to an oxide, indicating that reduction of HAuCl_4 precursor was complete within the detection limits of this technique.

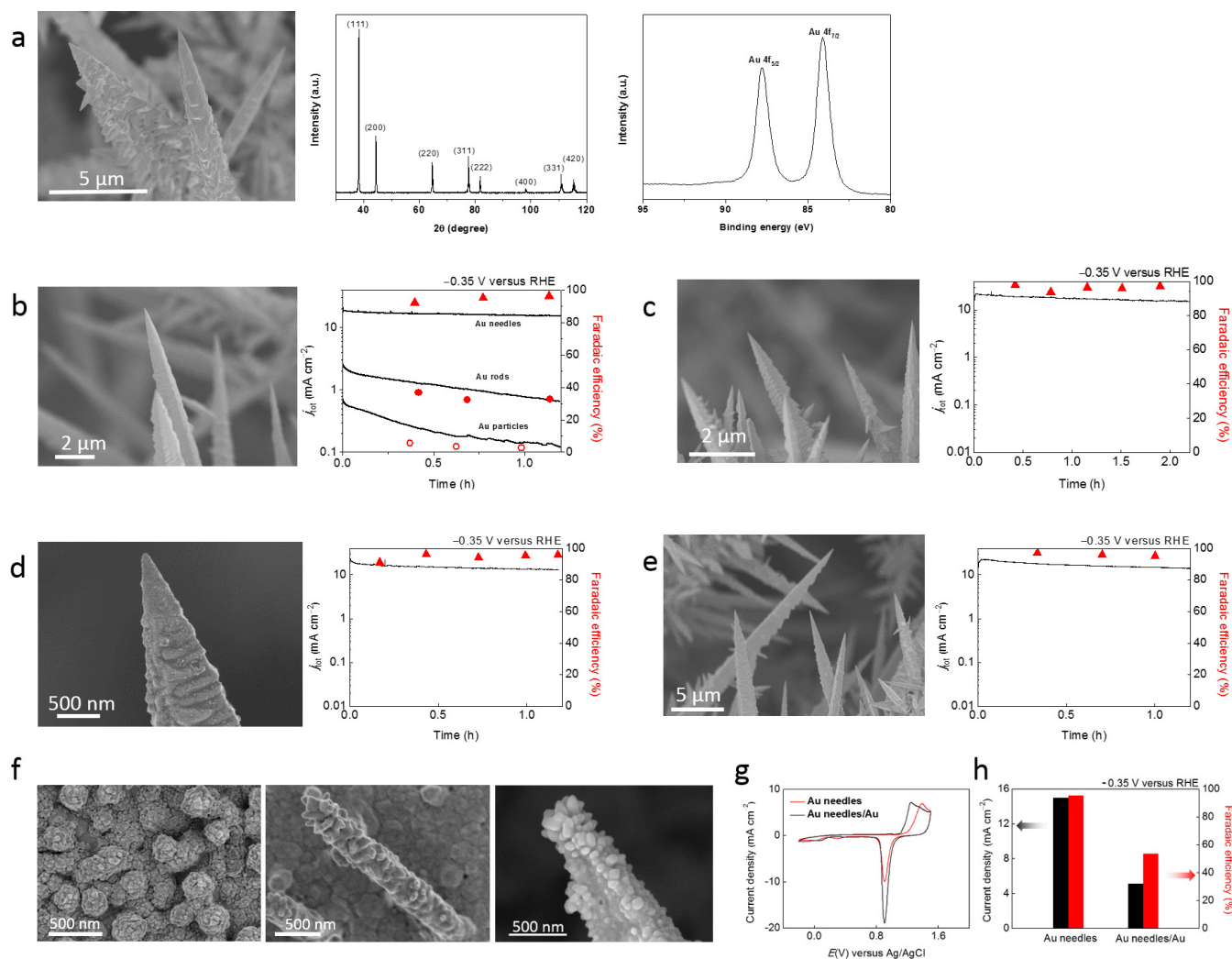
f, Current–voltage curves on the tips of single Au needle, rod and particle. The radii for the Au needle, rod and particle are 5 nm, 60 nm and 140 nm, respectively. **g**, Charge transfer resistance analyses. Nyquist plots in 0.5 M KHCO_3 aqueous electrolyte. **h**, **i**, CO_2 reduction performances in 0.5 M KHCO_3 , pH 7.2 at -0.30 V (**h**) and -0.20 V (**i**) versus RHE. **j**, **k**, CO_2 reduction current densities in 0.5 M KHCO_3 , pH 7.2, normalized by (j) geometric area and (k) ECSA. **l–n**, Activation energy analyses. The polarization curves of Au particles (**l**), Au rods (**m**), and Au needles (**n**) in 0.5 M KHCO_3 aqueous electrolyte at 0–25 $^\circ\text{C}$. Insets are the Arrhenius plots for the dependence of reaction rate for CO_2 reduction on temperature.



Extended Data Figure 4 | Collective control experiments to confirm that the reactivity of Au nanoneedles cannot be simply explained by oxides or adatoms. **a, b**, O 2p core-level X-ray photoelectron spectroscopy spectra (**a**) and O K-edge X-ray absorption spectra (**b**) for Au needles, pure Au and oxidized Au needles. The O 2p core-level and O K-edge X-ray absorption spectra of Au needles are similar to those of pure Au and are different from that of oxidized Au needles, indicating the different Au states in Au needles and oxidized Au. **c**, High-resolution TEM image of Au needle tip, indicating that there is no obvious facet and adatoms. **d**, Electron energy loss spectroscopy (EELS) spectra on Au needle tip.

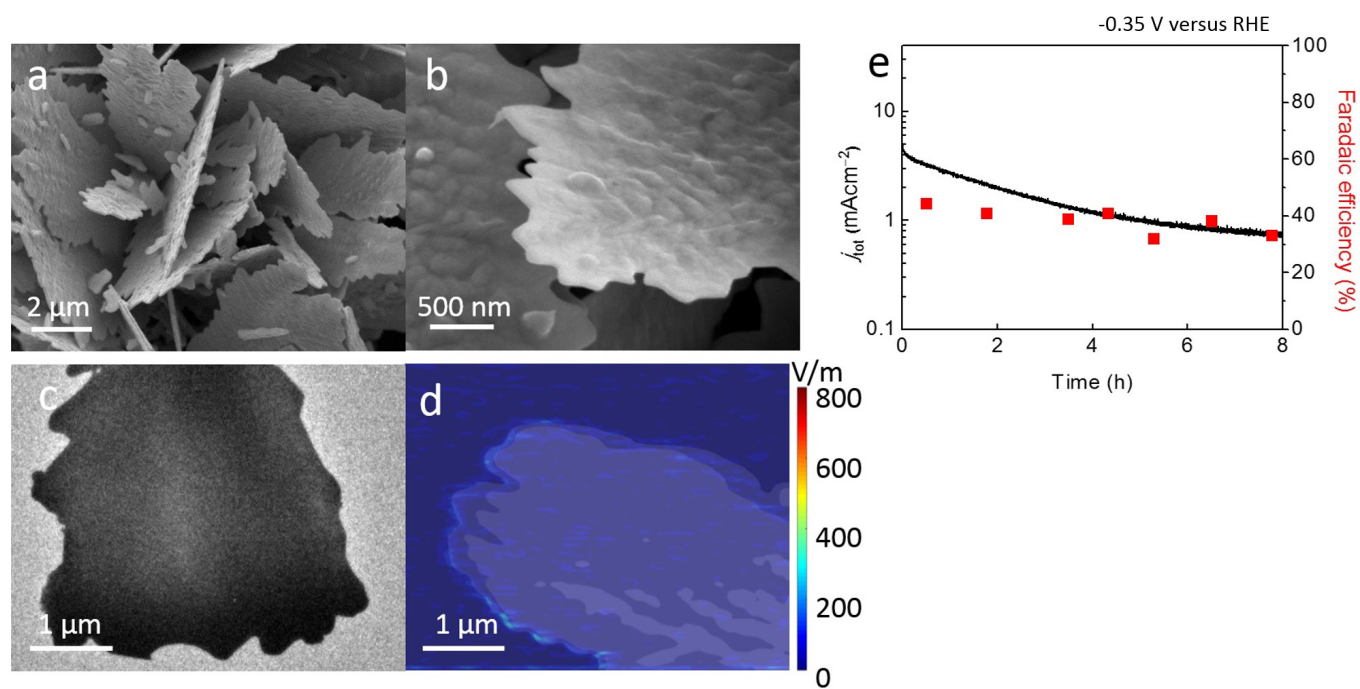


No oxide can be detected on Au needle tip, indicating that reduction of the HAuCl_4 precursor was complete within the detection limits of this technique. **e**, Low-magnification SEM image of oxidized Au needles. **f**, High-magnification SEM image of oxidized Au needles. **g**, TEM image of oxidized Au needles. Amorphous Au oxide can be observed on the surface of Au. **h**, X-ray photoelectron spectroscopy spectra of oxidized Au needles and primary Au needles. **i**, Cyclic voltammograms collected for Au needles and oxidized Au needles. **j**, CO_2RR performance on oxidized Au needles.

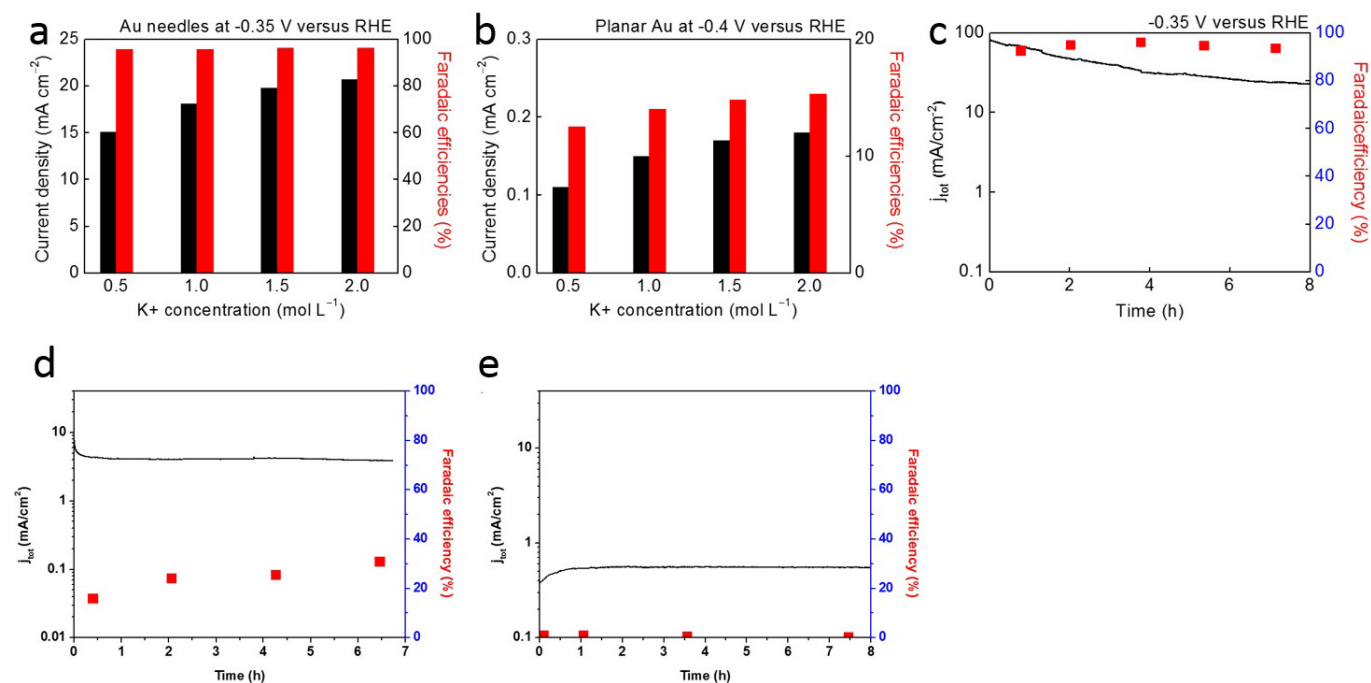


Extended Data Figure 5 | Collective control experiments to confirm the FIRC effects. **a**, Morphology, crystal structure and composition for Au needles after reaction. Left, SEM image, middle, X-ray diffraction pattern, and right, X-ray photoelectron spectroscopy spectrum for Au needles after long term CO₂RR. **b**, Left, SEM image of Au needles covered by 10-nm Au by electron beam deposition, right, CO₂ reduction activity of Au needles, rods and particles at -0.35 V versus RHE. **c**, Left, SEM image of Au needles at 140 °C after annealing, right, CO₂ reduction activity of Au needles at -0.35 V versus RHE after annealing. **d**, Left, SEM image of Au needles after surface etching. The Au nanoneedles were immersed in a vial containing 15 ml of CuCl₂ solution (5 mM). The vial was then heated

to 70 °C using an oil bath and kept at that temperature for 1 h. The etched Au nanoneedles obtained were washed with a copious amount of water and dried at room temperature²⁷. Right, CO₂ reduction activity of Au needles at -0.35 V versus RHE after surface etching. **e**, Left, SEM image of Au needles after surface plasma bombard (50 W, argon atmosphere, 1 h). Right, CO₂ reduction activity of Au needles at -0.35 V versus RHE after surface plasma bombard. **f**, SEM image of Au particles (left), Au rods (middle), and Au needles (right) with secondarily deposited Au particles. **g**, Cyclic voltammograms collected for Au needles in 50 mM H₂SO₄ for ECSA measurements. **h**, CO₂RR performances of Au needles and Au needles/Au at -0.35 V versus RHE.

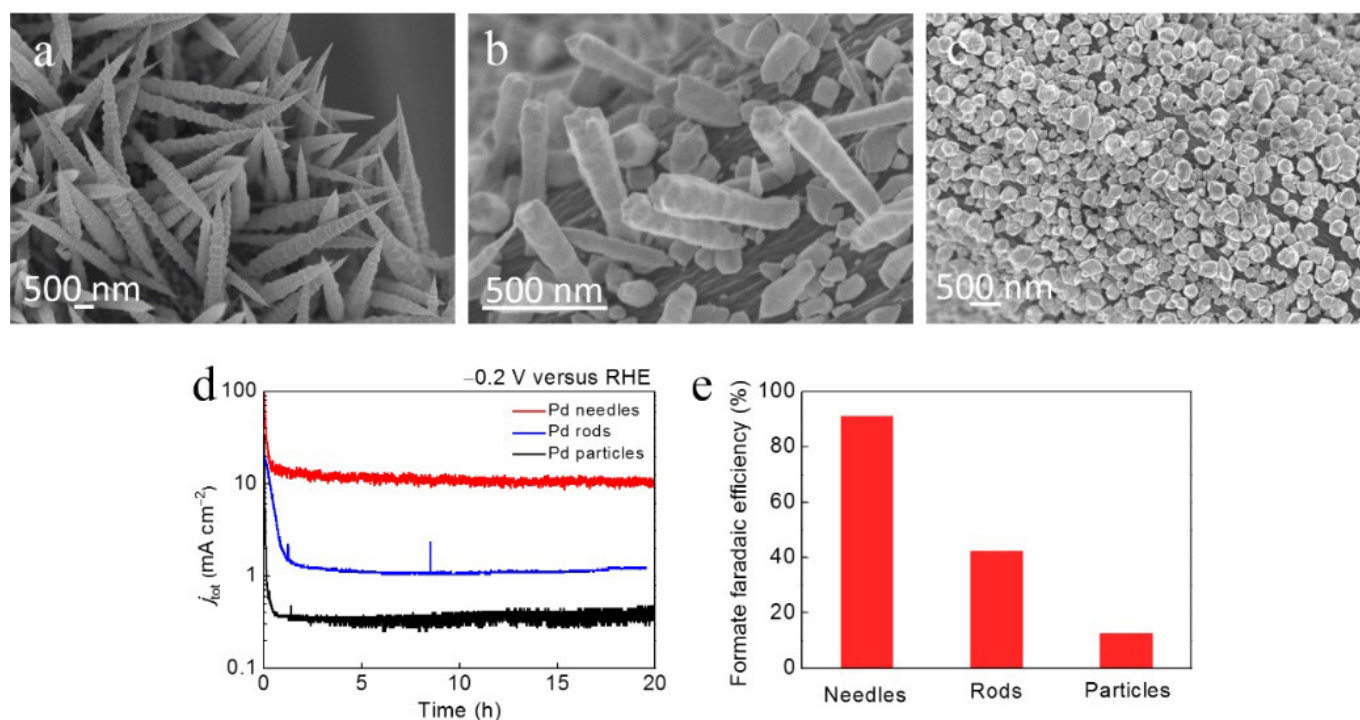


Extended Data Figure 6 | Morphology, electric field and CO₂ RR performance of dendritic Au leaves. **a, b,** SEM images of Au leaves. **c,** TEM image of Au leaves. **d,** Electric field distribution deduced using Kelvin probe atomic force microscopy. **e,** CO₂ reduction activity of Au leaves at -0.35 V versus RHE.



Extended Data Figure 7 | CO₂RR performances of Au nanoneedles in various electrolyte condition. **a**, Current densities and Faradaic efficiencies versus K⁺ concentrations on Au needles at -0.35 V versus RHE. **b**, Current densities and Faradaic efficiencies versus

K⁺ concentrations on planar Au at -0.4 V versus RHE. **c**, CO₂ reduction performance of Au needles in saturated KHCO₃ solution. **d**, CO₂ reduction performance of Au needles in NH₄HCO₃ solution. **e**, CO₂ reduction performance of Au needles in water.



Extended Data Figure 8 | CO₂ reduction reaction performances on Pd needles, rods and particles. **a–c**, SEM images of Pd needles, rods and particles, respectively. **d**, Total current density versus time for CO₂ RR on Pd needles, rods and particles in 0.5 M KHCO₃ solution at -0.2 V versus RHE. **e**, Average Faradaic efficiency for formate production versus time on Pd needles, rods and particles in 0.5 M KHCO₃ solution at -0.2 V versus RHE.

Extended Data Table 1 | Summary of simulation parameters as calculated from DFT

A) Free Energy Corrections for gas-phase species (eV)											
Species	E _{DFT}	ZPE	∫C _{vdT}	TΔS	G						
H ₂ O	-14.214	0.564	0.806	0.67	-14.218						
CO ₂	-22.946	0.306	0.099	0.662	-23.204						
H ₂	-6.771	0.268	0.091	0.434	-6.848						
CO	-14.775	0.132	0.091	0.668	-15.221						
HCOOH	-29.87	0.891	0.348	-1.047	-29.914						
B) Free Energies for CO ₂ RR Reaction (eV)											
Surface	Rxn Coordinate	ΔE									
		Au(111)	Au(100)	Au(110)	Au(211)						
Bare	CO ₂ + * + 2(H ⁺ + e ⁻)	0.00	0.00	0.00	0.00						
	COOH* + H ⁺ + e ⁻	1.50	1.28	0.97	0.92						
	CO* + H ₂ O	0.93	0.74	0.29	0.34						
	CO + * + H ₂ O	0.61	0.61	0.61	0.61						
K+	CO ₂ + * + 2(H+ + e-)	0.00	0.00	0.00	0.00						
	COOH* + H+ + e-	0.61	0.61	0.28	0.40						
	CO* + H ₂ O	0.69	0.71	0.18	0.42						
	CO + * + H ₂ O	0.61	0.61	0.61	0.61						
C) Free Energy Corrections for Surfaces and Adsorbates											
Facet	Surface	Species	E _{elec}	ZPE	∫C _{vdT}	TΔS	G				
Au(111)	Bare	*	-81.61								
		CO*	-96.53	0.17	0.10	-0.25	-96.51				
		COOH*	-107.27	0.65	0.09	-0.20	-106.73				
	K+	*	-84.07								
		CO*	-99.26	0.17	0.09	-0.21	-99.21				
		COOH*	-110.57	0.62	0.10	-0.24	-110.09				
Au(100)	Bare	*	-79.01								
		CO*	-94.25	0.20	0.07	-0.13	-94.11				
		COOH*	-104.88	0.64	0.10	-0.22	-104.37				
	K+	*	-81.49								
		CO*	-96.72	0.18	0.08	-0.15	-96.61				
		COOH*	-108.01	0.62	0.10	-0.22	-107.50				
Au(110)	Bare	*	-103.85								
		CO*	-119.57	0.22	0.06	-0.10	-119.40				
		COOH*	-130.13	0.67	0.08	-0.13	-129.51				
	K+	*	-106.41								
		CO*	-122.24	0.22	0.06	-0.10	-122.06				
		COOH*	-133.33	0.64	0.09	-0.16	-132.76				
Au(211)	Bare	*	-80.61								
		CO*	-96.24	0.20	0.07	-0.13	-96.10				
		COOH*	-106.87	0.65	0.09	-0.19	-106.32				
	K+	*	-83.42								
		CO*	-98.93	0.18	0.08	-0.17	-98.84				
		COOH*	-110.19	0.64	0.10	-0.19	-109.65				
D) Average closest Au-CO ₂ distance (Å)											
Facet	Without K ⁺	With K ⁺	Facet	Without K ⁺	With K ⁺	Facet	Without K ⁺	With K ⁺	Facet	Without K ⁺	With K ⁺
(111)	3.25	2.75	(110)	3.24	2.75	(100)	3.28	2.79	(211)	3.29	2.80

ZPE, zero-point vibrational energy; $\int C_{\text{vdT}}$, heat capacity; T , temperature, ΔS , entropy; G , Gibbs energy.

Extended Data Table 2 | Summary of ECSA, activation energies and charge transfer resistances on different Au electrodes

Sample	Geometric area 1 (cm ²)	^a ECSA 1 (cm ²)	Roughness factor 1	Geometric area 2 (cm ²)	^b ECSA 2 (cm ²)	Roughness factor 2	^c Activation energies (kJ mol ⁻¹)	^c Charge transfer resistance (Ω)
Au needles	0.18	9.59	53.28	0.13	6.70	51.54	21	24
Au rods	0.21	6.91	32.90	0.23	7.86	34.17	44	92
Au particles	0.26	3.26	12.54	0.28	3.41	12.18	72	240

^aECSA 1, electrochemically active surface area, determined by integrating the oxide reduction peak area obtained from cyclic voltammogram.^bECSA 2, electrochemically active surface area, determined by measuring anodic stripping waves for underpotential-deposited Cu monolayers.^cMeasured in 0.5 M KHCO₃.

Extended Data Table 3 | Summary of CO₂RR performances on different Au and Pd electrodes in aqueous solution with inorganic electrolyte

Sample	Electrolyte	Product	Potential vs. RHE (mV)	j_{product} (mA cm ⁻²)	FE	Onset over- potential (mV)	Tafel slope (mV dec ⁻¹)	Reference
Au needles	0.5 M KHCO ₃	CO	-350	~15	95%	70	42	This work
Au needles	^a Sat. KHCO ₃	CO	-350	~22	95%	70	—	This work
Au rods	0.5 M KHCO ₃	CO	-350	~0.7	25%	190	80	This work
Au particles	0.5 M KHCO ₃	CO	-350	~0.1	3%	240	96	This work
Oxide- derived Au	0.5M NaHCO ₃	CO	-350	~2	96%	140	56	Reference (9)
Au nanowire	0.5 M KHCO ₃	CO	-350	~1.8 ^b	94%	90	—	Reference (25)
Au NP's	0.5 M NaHCO ₃	CO	-350	<0.02	63%	190	—	Reference (24)
Pd needles	0.5 M KHCO ₃	HCOOH	-200	~10	91%	—	—	This work
Pd rods	0.5 MKHCO ₃	HCOOH	-200	~0.5	42%	—	—	This work
Pd particles	0.5 M KHCO ₃	HCOOH	-200	~0.05	13%	—	—	This work
Partially oxidized Co	0.1 M Na ₂ SO ₄	HCOOH	-200	~3	45%	—	—	Reference (2)
Pd NP's	0.5 M KHCO ₃	HCOOH	-200	~1	50%	—	—	Reference (28)
Pd NP's	2.8 M NaHCO ₃	HCOOH	-200	~0.4	82%	—	—	Reference (28)

^aSat. KHCO₃, saturated KHCO₃ solution.^bThe unit of current density is A g⁻¹.

Data is taken from refs 2, 9, 24, 25 and 28. NP, nanoparticle; FE, Faradaic efficiency.

Catalytic enantioselective 1,6-conjugate additions of propargyl and allyl groups

Fanke Meng¹, Xiben Li¹, Sebastian Torker¹, Ying Shi¹, Xiao Shen¹ & Amir H. Hoveyda¹

Conjugate (or 1,4-) additions of carbanionic species to α,β -unsaturated carbonyl compounds are vital to research in organic and medicinal chemistry, and there are several chiral catalysts that facilitate the catalytic enantioselective additions of nucleophiles to enoates¹. Nonetheless, catalytic enantioselective 1,6-conjugate additions are uncommon, and ones that incorporate readily functionalizable moieties, such as propargyl or allyl groups, into acyclic $\alpha,\beta,\gamma,\delta$ -doubly unsaturated acceptors are unknown². Chemical transformations that could generate a new bond at the C6 position of a dienolate are particularly desirable because the resulting products could then be subjected to further modifications. However, such reactions, especially when dienolates contain two equally substituted olefins, are scarce³ and are confined to reactions promoted by a phosphine-copper catalyst (with an alkyl Grignard reagent^{4,5}, dialkylzinc or trialkylaluminium compounds^{6,7}), a diene-iridium catalyst (with arylboroxines)^{8,9}, or a bisphosphine-cobalt catalyst (with monosilyl-acetylenes)¹⁰. 1,6-Conjugate additions are otherwise limited to substrates where there is full substitution at the C4 position¹¹. It is unclear why certain catalysts favour bond formation at C6, and—although there are a small number of catalytic enantioselective conjugate allyl additions^{12–15}—related 1,6-additions and processes involving a propargyl unit are non-existent. Here we show that an easily accessible organocopper catalyst can promote 1,6-conjugate additions of propargyl and 2-boryl-substituted allyl groups to acyclic dienolates with high selectivity. A commercially available allenyl-boron compound or a monosubstituted allene may be used. Products can be obtained in up to 83 per cent yield, >98:2 diastereomeric ratio (for allyl additions) and 99:1 enantiomeric ratio. We elucidate the mechanistic details, including the origins of high site selectivity (1,6- versus 1,4-) and enantioselectivity as a function of the catalyst structure and reaction type, by means of density functional theory calculations. The utility of the approach is highlighted by an application towards enantioselective synthesis of the anti-HIV agent (–)-equisetin.

Designing an efficient 1,6-conjugate addition is difficult in that the largest coefficient of the lowest unoccupied molecular orbital (LUMO; see compound **i** in Fig. 1a) is at the fourth carbon position (C4) and consequently this is where the bond is preferentially generated. We surmised that any conjugate addition should begin by interaction of the nucleophilic allenyl-copper species with dienolate C4 by an oxidative addition to give compound **ii** or π -complexation¹⁶ to afford product **iii** (Fig. 1a). 1,6-Addition could then be favourable if certain alternative processes were faster than the 1,1'-reductive elimination that affords the 1,4-allenyl addition compound **iv** (route A in Fig. 1a).

We envisioned two possible scenarios. (1) Compound **ii** might undergo a 1,3-shift followed by 1,1'-reductive elimination to deliver the 1,6-allenyl addition product **vi** via **v** by forming a bond between the C6 of the dienolate and the C α of the allenyl-metal system (route B); this type of π -allyl isomerization has been previously suggested⁵ but experimental or computational support has not been forthcoming. (2) Organocopper complex **iii** might be directly transformed to the

1,6-propargyl-addition product **vii** by formation of a bond between the dienolate C6 and the C γ of the allenyl-copper moiety (route C in Fig. 1a). That is, the allenyl-copper complex in its bent form (see Fig. 1c) may interact with the C3–C4 π cloud, placing the nucleophilic C γ near the dienolate C6. This pathway would be reminiscent of a 3,3'-reductive elimination proposed vis-à-vis enantioselective allyl-allyl coupling with Ni and Pd complexes^{17,18}. A similar reaction mode with an organocopper species has been mentioned in just one instance (again, without experimental or computational support)¹⁹.

We first carried out a model transformation involving the dienolate **1a** and the commercially available allenyl-B(pin) **2** (where pin is pinacolato) with a complex derived from imidazolinium salt **3a** and CuCl. We opted for NaOPh versus an alkoxide as the stoichiometric base (for example, NaO*tert*-Bu) because the residual CuOPh is less Lewis basic and would not interfere with the function of a chiral catalyst; small amounts of either may, however, be used to deprotonate the imidazolinium salt to generate the N-heterocyclic carbene (NHC)-Cu complex. In the event, at ambient temperature and after 16 h **4a** was isolated in 64% yield as a single alkene isomer (>98% β,γ -enoate); the 1,6-allenyl, 1,4-propargyl or 1,4-allenyl addition products were not detected (**5a–7a**). The exclusive formation of **4a** implies that the pathway involving a π -allyl shift is not operative (route B); otherwise, allenyl compound **6a** would be formed. The catalytic cycle in Fig. 1c is probably the most relevant. Under the same conditions but with the corresponding $\alpha,\beta,\gamma,\delta$ -unsaturated monoester there was minimal conversion (<5%).

Next, we examined the effect of chiral phosphorus-based ligands. In certain cases a complicated mixture of compounds was generated (**8a** and **8d**, Fig. 2a) and in others, unlike the aforementioned NHC copper species, appreciable amounts of the 1,4-addition product (**5a**) were formed (mechanistic analysis below). Enantioselectivity was uniformly low. Matters improved with NHC-Cu complexes (**9a–9e**, Fig. 2a): **4a** was generated exclusively (>98:2 1,6-:1,4-propargyl addition), and the complex generated from phenylglycine-derived **9b** delivered it in 74% yield and 97.5:2.5 enantiomeric ratio. Nevertheless, some of the screening data were unexpected. Whereas reaction with the larger **9c** afforded substantially reduced selectivity (67.5:32.5 enantiomeric ratio), with imidazolinium salt **9d**, which contains a less imposing 3,5-dimethylphenyl moiety, **4a** was formed in 92:8 enantiomeric ratio. Further, unlike the related catalytic allylic substitution processes²⁰, protection of the NHC hydroxy group was detrimental to enantioselectivity: reaction with silyl-protected **9e** gave nearly racemic product.

The enantioselective protocol has considerable scope (Fig. 2b). Dienolates with an aryl unit (**4b–4g**), whether it is electron-donating (**4e**) or electron-withdrawing (**4g**), react efficiently to give products in >98:2 propargyl:allenyl and 1,6-:1,4 selectivity and 94:6–97:3 enantiomeric ratio. A bromoaryl group is tolerated (**4f**), which is notable since an unhindered aryl-bromine bond can be prone to undergoing oxidative insertion with a copper(I) complex. High efficiency and selectivity was observed with heterocyclic substrates (**4h**, **4i**) or those

¹Department of Chemistry, Merkert Chemistry Center, Boston College, Chestnut Hill, Massachusetts 02467, USA.

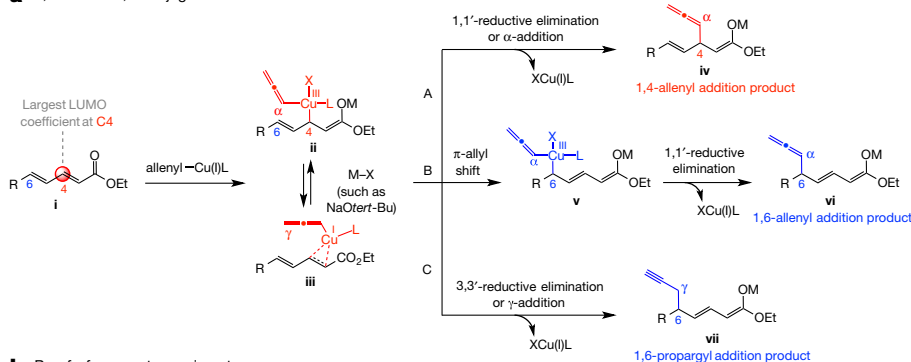
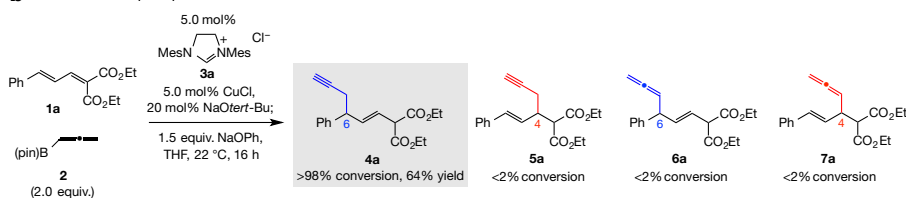
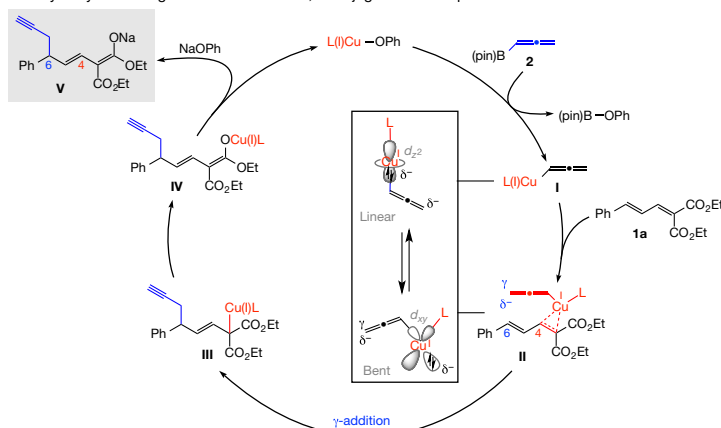
a 1,4- versus 1,6-conjugate addition reactions**b** Proof-of-concept experiment**c** Possible catalytic cycle leading to the formation of 1,6-conjugate addition products

Figure 1 | Possible conjugate addition pathways, the initial experiment and a plausible catalytic cycle. **a**, In a conjugate reaction, addition to the C4 site is kinetically favoured (\rightarrow **ii** or **iii**); subsequent 1,1'-reductive elimination (or α -addition) could afford product **iv** (route A), or a 1,3- π -allyl shift (\rightarrow **v**) may precede reductive elimination, affording 1,6-allenyl addition product **vi** (route B). Alternatively, **ii** or **iii** may be

directly converted to **vii** by a γ -addition (3,3'-reductive elimination type) process (route C). **b**, Proof-of-principle experiment indicates that with an allenyl-copper intermediate, route C predominates. **c**, Plausible catalytic cycle for the preferential formation of the 1,6-propargyl addition product. R and G, various organic functional groups; LUMO, lowest unoccupied molecular orbital; M, metal; Mes, 2,4,6-trimethylphenyl.

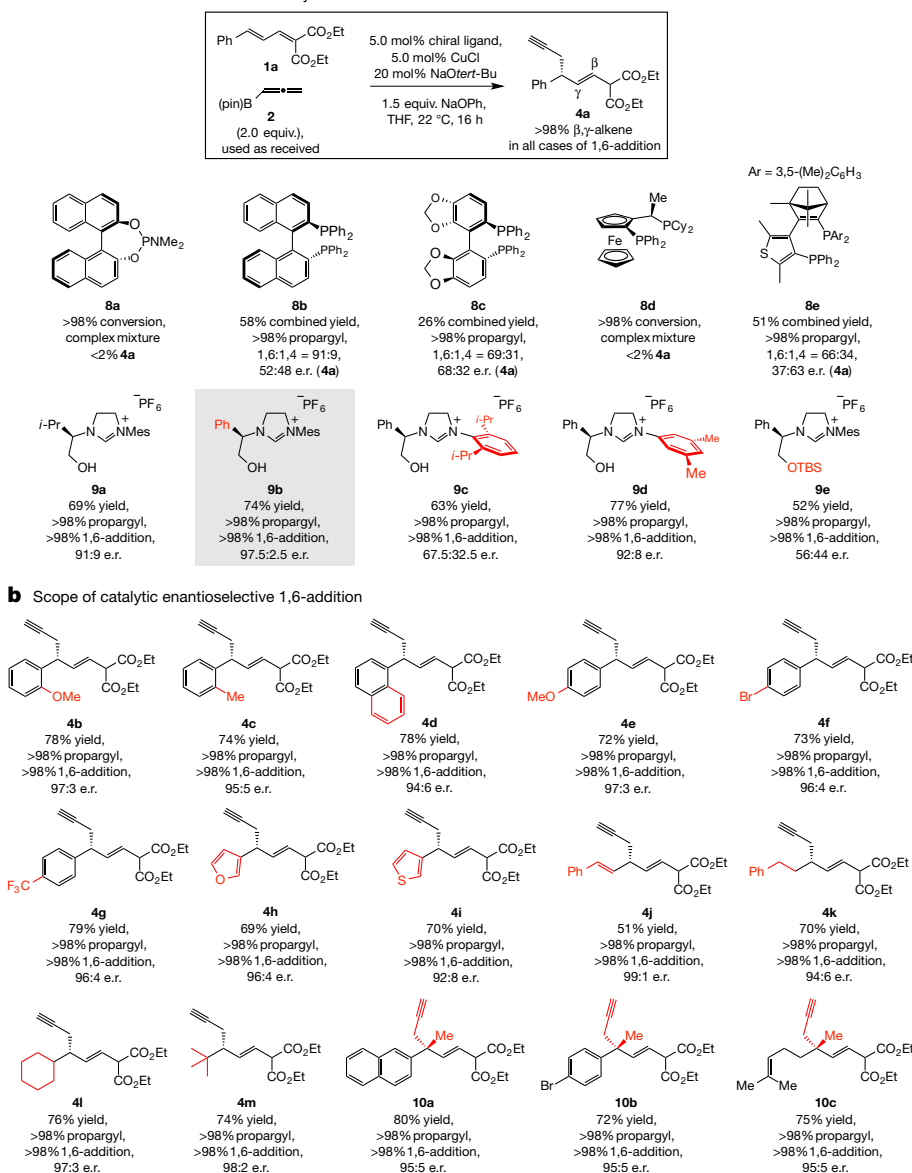
that bear an allenyl (**4j**), a linear or a branched aliphatic group (**4k**, **4l**). The sterically congested *tert*-butyl-substituted **4m** was isolated in 74% yield and 98:2 enantiomeric ratio (>98% propargyl and 1,6-addition); the ease with which this C–C bond is generated confirms that the initial addition occurs at C4, distally from the quaternary site at C6, followed by an intramolecular event that is less susceptible to steric pressure. All-carbon quaternary stereogenic centres²¹ were formed efficiently and with exceptional group-, site- and enantioselectivity (**10a–10c**, Fig. 2b). Again, there were no allenyl- or 1,4-addition byproducts and a single olefin isomer (>98% *E*) was detected (400 MHz ¹H NMR analysis; see Supplementary Information for details.)

We then evaluated the possibility of a multicomponent enantioselective 1,6-addition by which a dienophile, a monosubstituted allene and B₂(pin)₂ may be combined (Fig. 3a). We envisioned association of allylcopper species **VI** with a substrate to yield **VII**, which could rearrange to give **VIII**. Electronic modification of one of the reacting alkenes by a B(pin) moiety could counter γ -addition. The other concern was that diastereomeric mixtures could form.

Reaction of dienophile **1a**, allene **11** and B₂(pin)₂ with 5.0 mol% imidazolium salt **3a** and CuCl afforded **12a** exclusively and with exceptionally high 1,6-:1,4- and diastereomeric ratios (Fig. 3b). The

low yield of **12a** arises from a breakdown in chemoselectivity, namely by competitive boryl 1,4-addition. The optimal catalyst would therefore have to deliver high enantioselectivity and favour Cu–B addition to the allene over its reaction with a dienophile. Examination of different chiral imidazolium salts led to encouraging results, as **12a** was generated more efficiently (57%–74% yield). The Cu complex derived from imidazolium salt **9d** proved optimal, affording **12a** in 62% yield, >98:2 diastereomeric ratio and 95:5 enantiomeric ratio. However, the trends in enantioselectivity were again puzzling. It was not a surprise that the transformations with catalysts derived from imidazolium salts **9b** and **9c** gave **12a** in 89:11 and 75:25 enantiomeric ratios, respectively, as there was a similar trend (albeit with a larger difference) with the propargyl additions (see Fig. 2a). What was perplexing was that enantioselectivity was higher with the less sterically congested **9d** (95:5 enantiomeric ratio).

Use of a diboryl compound such as **13** (ref. 22; Fig. 3c) is a less attractive option. The need for the initial synthesis of an allylboron compound notwithstanding, the two-stage alternative, although highly γ -, site- and diastereoselective, proceeds with diminished enantioselectivity (Fig. 3c): with **13** as the reagent, **12a** was obtained in 81:19 enantiomeric ratio (versus 95:5 through the multicomponent process).

a Identification of an effective chiral catalyst**Figure 2 | Catalytic enantioselective 1,6-propargyl conjugate additions.**

a, Screening of a variety of chiral phosphine and NHC ligands indicated that the chiral copper catalysts derived from the latter series are much more effective, and that corresponding to imidazolium salt **9b** is optimal. **b**, The catalytic process is broadly applicable, affording products uniformly with >98% propargyl and 1,6-addition selectivity and in up to 80% yield and 98:2 enantiomeric ratio (e.r.). Products containing a tertiary or an all-carbon quaternary carbon stereogenic centre can be accessed. TBS, *tert*-butyldimethylsilyl; THF, tetrahydrofuran. Reactions were performed

under N₂ under the conditions shown in the box for **4a**, except for **10a–10c**, where 10 mol% **9b** and CuCl were used and the mixture was allowed to stir for 24 h. Conversions, propargyl:allenyl and 1,6:1,4-addition ratios were measured by analysis of ¹H NMR spectra of unpurified mixtures; the variance of values estimated to be less than about 2%. Yields correspond to isolated and purified products and represent an average of at least three runs (±5%). See Supplementary Information for experimental details and spectroscopic analyses.

Control experiments indicate that lower enantiomeric ratio originates from efficient addition of an achiral allylcopper species, generated from allyl(PhO)CuNa with **13**, to the dienolate to give racemic product; the background reaction produces the 1,6-addition isomers exclusively (see below for further discussion and Supplementary Information for mechanistic or computational analysis). In the multicomponent reactions, on the other hand, only an NHC–Cu(OPh) complex can efficiently activate the B–B bond in B₂(pin)₂.

The three-component process has ample range as well (Fig. 3e). Aryl- (**12b–12f**), heteroaryl- (**12g**, **12h**), or alkyl-substituted (**12i**) dienolates can be converted to the desired products with >98% γ-, site-, and diastereoselectivity and 93:7–99:1 enantiomeric ratio. Compounds containing an alkene (**12j**), an alkyne (**12k**) or a Weinreb amide (**12l**) were synthesized in 58–74% yield and 88:12–95:5

enantiomeric ratio, and an unsubstituted allene may be used (**12m**). Unlike with allenylboronate **2** (see Fig. 2b), reactions with the more highly substituted enolates that would afford quaternary carbon centres were inefficient (<10% conversion); this is probably because of the more severe steric repulsion caused by the sizeable B(pin) moiety.

Several key mechanistic questions needed to be addressed at this point: why does the identity of the optimal chiral Cu complex vary so much, and why is it that, unlike multicomponent allylic substitutions²⁰, the unprotected hydroxyl group within the catalyst structure is needed for high enantioselectivity? To shed light on these issues, we performed density functional theory (DFT) calculations at the ωB97XD/Def2TZVP//ωB97XD/Def2SVP_{THF(PCM)} level of theory (see Supplementary Information for details). Computational studies (Fig. 4) indicate that the linear Cu(I)–allenyl species derived from

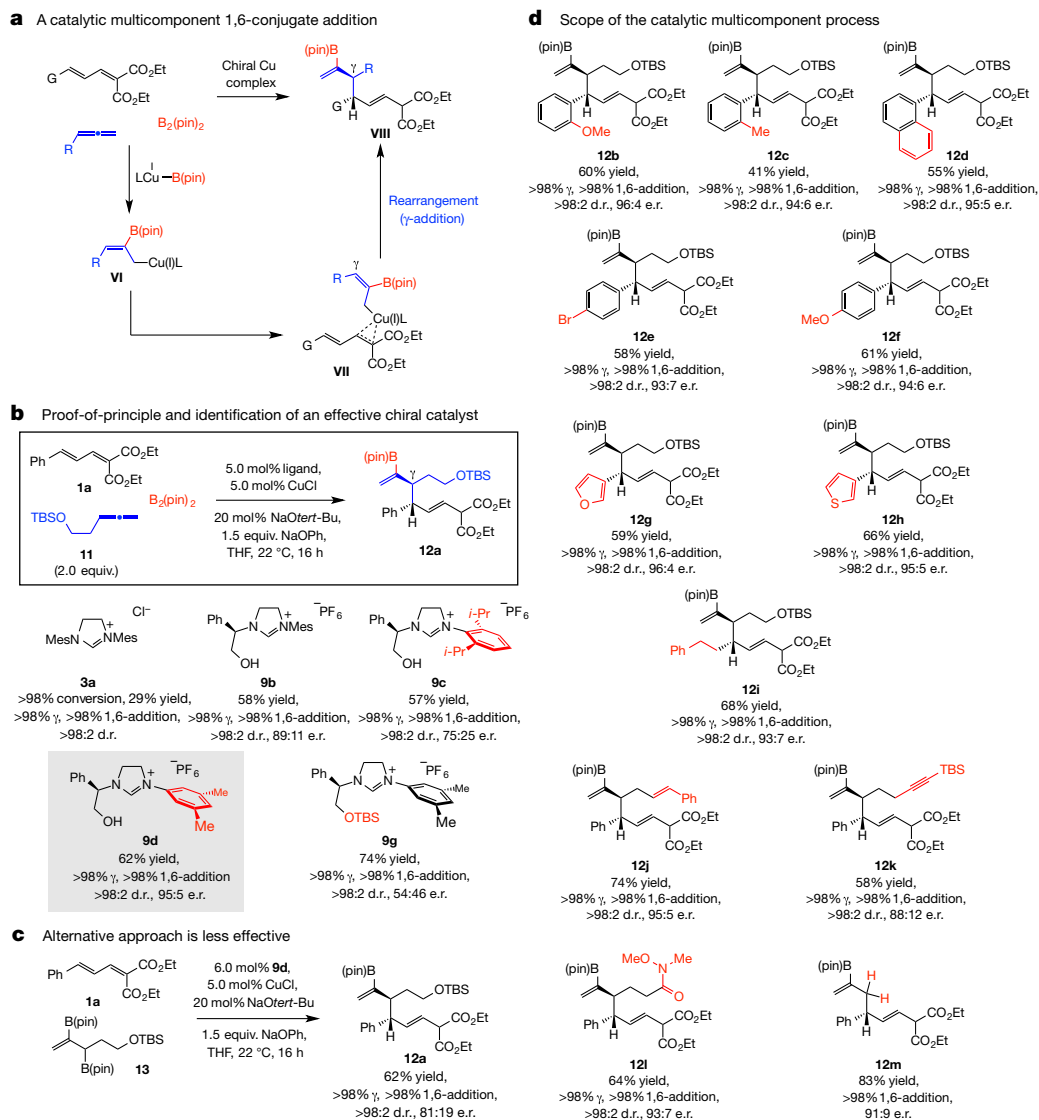


Figure 3 | Catalytic diastereo- and enantioselective multicomponent 1,6-conjugate addition of 2-B(pin)-substituted allyl moieties. **a**, The pathway through which 1,6-addition products may be generated by a multicomponent process involving a dienoate, an allene and $B_2(\text{pin})_2$. **b**, Preliminary experiment with an achiral NHC–Cu complex demonstrates that, although inefficient, reactions are exceptionally γ -, group- and 1,6-selective. Screening studies to identify an effective chiral catalyst indicates that a different NHC ligand is optimal for these transformations (versus propargyl additions). TBS, *tert*-butyldimethylsilyl. **c**, The alternative approach entailing initial synthesis of a diboryl reagent

imidazolium salts **9b** and **9d** associate with the C3–C4 bond, furnishing a square planar-type π -complex¹⁶ (**A–D**, Fig. 4a), precursors to 1,6-propargyl addition products (Fig. 2). The propensity of NHC–Cu complexes to afford an η^2 complex at the C3–C4 site may be attributed to the stronger electron-donating ability of the heterocyclic ligands (versus phosphine), which raises the energy of copper's *d*-orbitals, causing stronger binding with the corresponding π^* orbital^{16,23}. 1,6-Addition products can be formed once an NHC–Cu– π complex is assembled. A relevant experimental finding is that, similar to NHC–Cu but unlike phosphine Cu complexes, reactions with allenyl(*tert*-BuO)CuNa (generated without a ligand) proceed with exceptional 1,6:1,4-selectivity (>98:2; see Supplementary Information for detailed analysis). This might be because, as with an NHC–Cu system, the strongly π -basic *in situ*-formed (allenyl)₂cuprate species [from allenyl(*tert*-BuO)CuNa] can establish a complex with the C3–C4 alkene (see **viii**→**ix** in Fig. 4b), resulting in 1,6-propargyl

addition. With the less Lewis basic phosphines, back-bonding is less favoured. Additionally, DFT calculations indicate that with a weaker σ -donating phosphine unit, the highest occupied molecular orbital (HOMO; dz^2 , Fig. 1c) in a linear allenyl–Cu–phosphine complex is lower in energy (–7.43 eV for L = PPh₃ versus –7.14 eV for L = NHC). Consequently, aryloxide–Cu complexation can be stronger (less ‘filled–filled’ electronic repulsion) during **x**→**xi** (Fig. 4b), leading to 1,4-propargyl addition. It follows that transformations proceeding via these conformationally more flexible transition structures are less enantioselective (versus NHC–Cu systems; see Fig. 2a).

DFT calculations reveal that high enantioselectivity originates from the structural organization caused by a cationic sodium interacting with the catalyst's hydroxyl unit, in turn H-bonded with the phenoxy counterion and the dienoate carbonyl groups. In the more favourable pathway with **9b** (via **A**; Fig. 4a, left panel) there is less steric strain between the N-aryl's methyl group and the allenyl hydrogen

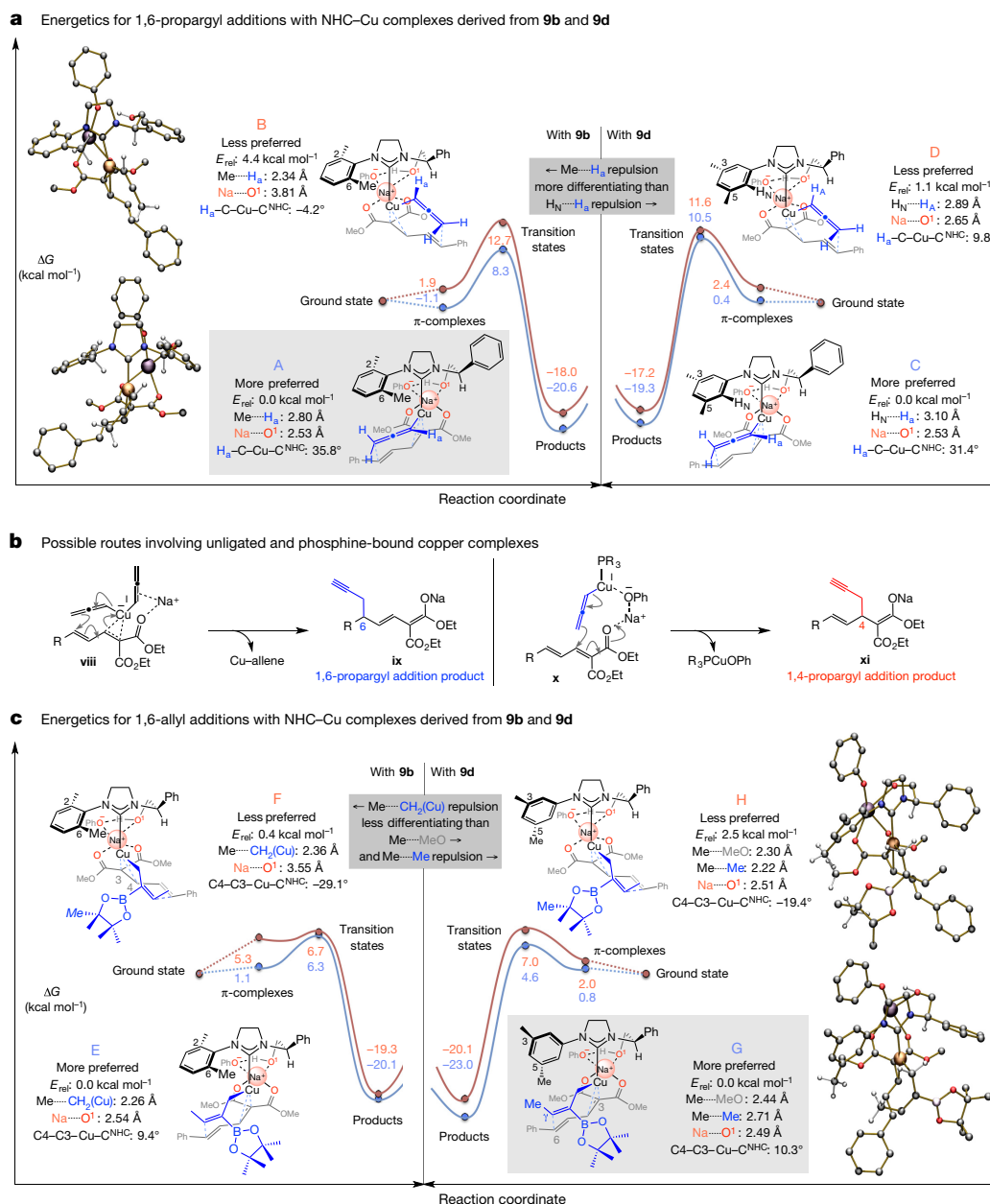


Figure 4 | Mechanistic considerations. **a**, Based on DFT calculations [ω B97XD/Def2TZVP// ω B97XD/Def2SVP level of theory (with THF as the solvent)] stereochemical models were developed for NHC–Cu-catalysed 1,6-propargyl additions with catalysts bearing an N-mesityl moiety (from **9b**). The issue is the larger energetic differentiation arising from steric repulsion between an *ortho*-methyl unit of the N-aryl group and the allenyl–copper moiety (that is, $\text{Me}^{\text{---}}\text{H}_a$, **9b**) versus one involving an aryl proton (that is, $\text{H}_N^{\text{---}}\text{H}_a$, **9d**). **b**, Routes by which a phosphine-

H_a than in **B** ($\text{Me}^{\text{---}}\text{H}_a$ distance of 2.80 Å versus 2.34 Å, respectively). Steric pressure in **B** can be alleviated by rotation of the NHC ligand around the Cu–C^{NHC} bond but this weakens the interaction between the hydroxy and the sodium cation (that is, $\text{Na}^{\text{---}}\text{O}^1$, 3.81 versus 2.53 Å in **B** and **A**, respectively). With the complex derived from the 3,5-dimethyl-substituted **9d** (Fig. 4a, right panel) a smaller energy gap separates the two modes of reaction (**C** versus **D**; 1.1 kcal mol^{−1} versus 4.4 kcal mol^{−1} for **A** and **B**); this might be because there is less difference in steric repulsion between the allenyl hydrogen and the *ortho* hydrogen of the N-aryl group in **D** ($\text{H}_N^{\text{---}}\text{H}_a$, 3.10 Å and 2.89 Å in **C** and **D**, respectively).

With the larger pin(B)-substituted allylcopper species (**E–H**, Fig. 4c) a similar complexation involving the catalyst's hydroxy unit and

based and non-ligated Cu complex might generate products, respectively. **c**, Transition state energies for enantioselective allyl additions are consistent with the observation that the catalyst derived from **9d** is optimal (versus **9b**). Steric repulsion involving a *meta*-methyl group of the NHC ligand with the carboxylic ester and the allylcopper substituents are the distinguishing elements. See Supplementary Information for details of calculations. NHC, N-heterocyclic carbene; E_{rel} , relative energy; ΔG , change in Gibbs free energy.

a sodium cation takes hold. The smaller NHC–Cu system (from the 3,5-dimethyl-phenyl-substituted **9d**) can differentiate better between the two orientations of the allylic nucleophile compared to when **9b** is involved (Fig. 4b, right panel). Specifically, in the transition state leading to the major enantiomer **G** the allylcopper moiety's C γ is oriented such that there is optimal overlap with C6 of the dienophile. In complex **H**, alignment of the latter two components engenders distortion of the C4–C3–Cu–C^{NHC} dihedral angle (−19.4° versus +10.3° for **H** and **G**, respectively). The N-aryl methyl unit in **H** sits closer to the bulky pinacolato moiety (2.22 Å) and a methoxy group (2.30 Å) than in **G** when it is at a more favourable distance from the allylic methyl and ester substituents (2.71 Å and 2.44 Å, respectively). With the less selective catalyst derived from **9b**, which contains an

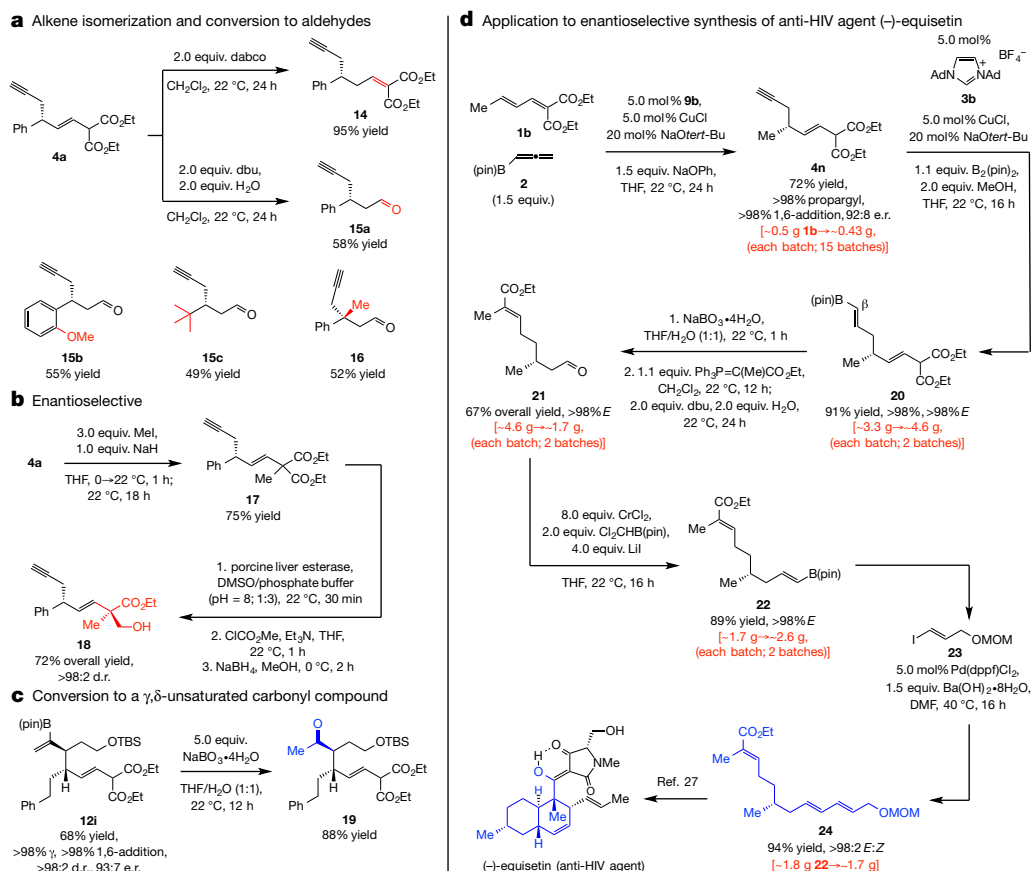


Figure 5 | Functionalizations and demonstration of utility.

a, The kinetically favoured alkene can be readily isomerized to the thermodynamically preferred isomer under one set of basic conditions, whereas in the presence of water, diester cleavage leads to the formation of β -substituted aldehydes. **b**, Alkylation followed by enzymatic desymmetrization of the diester unit proceeds with excellent stereochemical control. **c**, Oxidation of the alkenyl–B(pin) moiety affords otherwise difficult-to-access γ,δ -unsaturated ketones with vicinal

N-2,6-dimethylphenyl group (Fig. 4b, left panel), the steric repulsion involving the N-aryl methyl and the methylene unit of the allylcopper moiety is the dominant interaction in **F** as well as **E** (Me^{CH_2} , 2.26 Å and 2.36 Å in **E** and **F**, respectively). Hence, overall, the calculated energy difference between **E** and **F** amounts to just 0.4 kcal mol^{–1} (versus 2.5 kcal mol^{–1} for **H** versus **G**; Fig. 4c).

The enantiomerically enriched products can be converted to otherwise difficult-to-access molecules (Fig. 5). Synthesis of enyne **14** (95% yield, Fig. 5a) shows that the kinetically generated β,γ -alkene can be equilibrated to the lower-energy, conjugated isomer. In contrast, with an amine base and water, the corresponding β -substituted aldehydes were formed exclusively: **15a–15c** and **16**, containing a quaternary carbon stereogenic centre, were obtained in 49%–58% yield. These transformations offer an attractive entry for synthesis of β -substituted enantiomerically enriched aldehydes that cannot be prepared easily by alternative protocols; as already mentioned, there are no catalytic enantioselective 1,4-propargyl additions¹. Synthesis of this type of enantiomerically enriched aldehydes by conjugate addition of an aryl- or alkyl-metal reagent (for example, PhMgCl or Me_2Zn) to an unsaturated ester followed by oxidation state adjustment would be problematic owing to the sensitivity of the methylene unit in the $\alpha,\beta,\delta,\omega$ -dienoate (or the corresponding enyne). The diester group offers additional opportunities; for example, through an alkylation/enzymatic desymmetrization²⁴ sequence, **4a** was converted to alcohol-ester **18** (Fig. 5b), which bears a quaternary carbon stereogenic centre. As represented by **19** (Fig. 5c), oxidation of the alkenyl–B(pin) moiety affords γ,δ -unsaturated ketones with vicinal stereogenic centres;

stereogenic centres at the α - and β -carbon sites. **d**, Application to synthesis of gram quantities of enantiomerically enriched triene **24**, previously used²⁷ in the total synthesis of the anti-HIV agent (–)-equisetin showcases utility of the catalytic approach. dabco, 1,4-diazabicyclo[2.2.2]octane; dbu, 1,8-diazabicyclo[5.4.0]undec-7-ene; DMSO, dimethylsulfoxide; Ad, adamantyl; dppe, 1,1'-bis(diphenylphosphino)ferrocene; MOM, methoxymethyl.

these fragments have been used in total synthesis of biologically active compounds and might be prepared by Claisen rearrangement²⁵ (see Supplementary Information for additional references) but are difficult to access directly, especially in a catalytic and enantioselective manner²⁶.

Preparation of triene **24**, employed in an intramolecular Diels–Alder reaction en route to the anti-HIV agent (–)-equisetin²⁷, highlights utility (Fig. 5c). Enyne **4n** was secured with >98% propargyl and 1,6-selectivity in 72% yield and 92:8 enantiomeric ratio. This transformation was performed at the 0.5 g scale (15 times) and with unpurified commercially available organoboron reagent **2**, underscoring the ease with which substantial quantities of the ligand precursor can be synthesized, the scalability of the catalytic processes and their high degree of reliability and reproducibility. NHC–Cu-catalysed site- and stereoselective proto-boryl addition²⁸ to the alkyne [($<2\%$ addition to alkene, $<2\%$ Z alkenyl–B(pin))] afforded **20** in 91% yield (about 4.6 g). Oxidation of the C–B bond, Wittig reaction ($>98\%$ E), followed by cleavage of the β,γ -unsaturated diester moiety to the derived aldehyde (see Fig. 5a), afforded **21** in 67% overall yield (about 1.7 g). The E-alkenyl–B(pin) **22** was then prepared in 89% yield (about 2.6 g) by a Cr(II)-based reagent (oxidation state with low toxicity)²⁹. Phosphine–Pd-catalysed cross-coupling with alkenyl iodide **23** (ref. 30) delivered approximately 1.7 g of the desired triene **24** (94% yield; $>98\%$ E). Although a pathway of similar length may be envisioned starting from readily available enantiomerically pure starting materials (for example, citronellol), this strategy has the advantage of being more easily amenable to analogue preparation.

We thus introduce an approach for efficient catalytic enantioselective 1,6-conjugate addition of two types of valuable unsaturated organic moieties to dienones. The mechanistic details regarding various features of an effective catalyst should help pave the way for achieving such important objectives. Considering that other readily available unsaturated organoboron or organocopper systems may be used, the low cost and ease with which the catalysts can be accessed, the reliably high selectivity values, and the versatility of the resulting products, the strategies presented here presage a considerable impact on future advances in stereoselective catalysis and chemical synthesis.

Received 22 April; accepted 21 June 2016.

Published online 1 August 2016.

- Alexakis, A., Krause, N. & Woodward, S. in *Copper-Catalyzed Asymmetric Synthesis* (eds Alexakis, A., Krause, N. & Woodward, S.) 33–68 (VCH–Wiley, 2014).
- Silva, E. M. P. & Silva, A. M. S. 1,6-Conjugate addition of nucleophiles to $\alpha,\beta,\gamma,\delta$ -diunsaturated systems. *Synthesis* **44**, 3109–3128 (2012).
- Tissot, M., Li, H. & Alexakis, A. in *Copper-Catalyzed Asymmetric Synthesis* (eds Alexakis, A., Krause, N. & Woodward, S.) 69–84 (VCH–Wiley, 2014).
- den Hartog, T., Harutyunyan, S. R., Font, D., Minnaard, A. J. & Feringa, B. L. Catalytic enantioselective 1,6-conjugate addition of Grignard reagents to linear dienones. *Angew. Chem. Int. Ed.* **47**, 398–401 (2008).
- den Hartog, T., van Dijken, D. J., Minnaard, A. J. & Feringa, B. L. An enantioselective approach to *syn* deoxypropionate units combining Cu-catalyzed 1,6- and 1,4-conjugate addition. *Tetrahedron Asymmetry* **21**, 1574–1584 (2010).
- Tissot, M. & Alexakis, A. Enantio- and regioselective conjugate addition of organometallic reagents to linear polyconjugated nitroolefins. *Chemistry* **19**, 11352–11363 (2013).
- Magrez-Chiquet, M. *et al.* Enantioselective 1,6-conjugate addition of dialkylzinc reagents to acyclic dienones catalyzed by Cu-DiPPAM complex—extension to asymmetric sequential 1,6/1,4-conjugate addition. *Chemistry* **19**, 13663–13667 (2013).
- Nishimura, T., Yasuhara, Y., Sawano, T. & Hayashi, T. Iridium/chiral diene-catalyzed asymmetric 1,6-addition of arylboroxines to $\alpha,\beta,\gamma,\delta$ -unsaturated carbonyl compounds. *J. Am. Chem. Soc.* **132**, 7872–7873 (2010).
- Nishimura, T., Noishiki, A. & Hayashi, T. Electronic tuning of chiral diene ligands in iridium-catalyzed asymmetric 1,6-addition of arylboroxines to γ -aryl- $\alpha,\beta,\gamma,\delta$ -unsaturated ketones. *Chem. Commun.* **48**, 973–975 (2012).
- Sawano, T., Ashouri, A., Nishimura, T. & Hayashi, T. Cobalt-catalyzed asymmetric 1,6-addition of (triisopropylsilyl)-acetylene to $\alpha,\beta,\gamma,\delta$ -unsaturated carbonyl compounds. *J. Am. Chem. Soc.* **134**, 18936–18939 (2012).
- Hénon, H., Mauduit, M. & Alexakis, A. Regiodivergent 1,4 versus 1,6 asymmetric copper-catalyzed conjugate addition. *Angew. Chem. Int. Ed.* **47**, 9122–9124 (2008).
- Sieber, J. D. & Morken, J. P. Asymmetric Ni-catalyzed allylation of activated enones. *J. Am. Chem. Soc.* **130**, 4978–4983 (2008).
- Shizuka, M. & Snapper, M. L. Catalytic enantioselective Hosomi–Sakurai conjugate allylation of cyclic unsaturated ketoesters. *Angew. Chem. Int. Ed.* **47**, 5049–5051 (2008).
- Kuang, Y. *et al.* Catalytic asymmetric conjugate allylation of coumarins. *Org. Lett.* **13**, 3814–3817 (2011).
- Yanagida, Y., Yazaki, R., Kumagai, N. & Shibasaki, M. Asymmetric synthesis of isothiazoles through Cu catalysis: direct catalytic asymmetric conjugate addition of allyl cyanide to α,β -unsaturated thioamides. *Angew. Chem. Int. Ed.* **50**, 7910–7914 (2011).
- Yoshikai, N. & Nakamura, E. Mechanisms of nucleophilic organocopper(I) reactions. *Chem. Rev.* **112**, 2339–2372 (2012).
- Keith, J. A. *et al.* The inner-sphere process in the enantioselective Tsuji allylation reaction with (S)-*t*-Bu-phosphinoxazoline ligands. *J. Am. Chem. Soc.* **129**, 11876–11877 (2007).
- Ardolino, M. J. & Morken, J. P. Congested C–C bonds by Pd-catalyzed enantioselective allyl–allyl cross-coupling, a mechanism-guided solution. *J. Am. Chem. Soc.* **136**, 7092–7100 (2014).
- Hornillos, V., Pérez, M., Fañanás-Mastral, M. & Feringa, B. L. Copper-catalyzed enantioselective allyl–allyl cross-coupling. *J. Am. Chem. Soc.* **135**, 2140–2143 (2013).
- Meng, F., McGrath, K. P. & Hoveyda, A. H. Multifunctional organoboron compounds for natural product synthesis. *Nature* **513**, 367–374 (2014).
- Quasdorf, K. E. & Overman, L. E. Catalytic enantioselective synthesis of quaternary carbon stereocenters. *Nature* **516**, 181–191 (2014).
- Pelz, N. F., Woodward, A. R., Burks, H. E., Sieber, J. D. & Morken, J. P. Palladium-catalyzed enantioselective diboration of prochiral allenes. *J. Am. Chem. Soc.* **126**, 16328–16329 (2004).
- Denmark, S. E. & Beutner, G. L. Lewis base catalysis in organic synthesis. *Angew. Chem. Int. Ed.* **47**, 1560–1638 (2008).
- Dominguez de María, P., García-Burgos, C. A., Bargeman, G. & van Gemert, R. W. Pig liver esterase (PLE) as biocatalyst in organic synthesis: From nature to cloning and to practical applications. *Synthesis* 1439–1452 (2007).
- Becker, J. *et al.* Total synthesis of (–)-ecklonialactone B. *Org. Lett.* **15**, 5982–5985 (2013).
- Krautwald, S., Schafroth, M. A., Sarlah, D. & Carreira, E. M. Stereodivergent α -allylation of linear aldehydes with dual iridium and amine catalysis. *J. Am. Chem. Soc.* **136**, 3020–3023 (2014).
- Yuki, K., Shindo, M. & Shishido, K. Enantioselective total synthesis of (–)-equisetin using a Me_3Al -mediated intramolecular Diels–Alder reaction. *Tetrahedron Lett.* **42**, 2517–2519 (2001).
- Jang, H., Zhugralin, A. R., Lee, Y. & Hoveyda, A. H. Highly selective methods for synthesis of internal (α -) vinylboronates through efficient NHC–Cu-catalyzed hydroboration of terminal alkynes. Utility in chemical synthesis and mechanistic basis for selectivity. *J. Am. Chem. Soc.* **133**, 7859–7871 (2011).
- Takai, K. *et al.* Transformation of aldehydes into (E)-1-alkenylboronic esters with a geminal dichromium reagent derived from a dichloromethylboronic ester and CrCl_2 . *Synlett* 963–964 (1995).
- Johansson Seechurn, C. C. C., Kitching, M. O., Colacot, T. J. & Sniekus, V. Palladium-catalyzed cross-coupling: a historical contextual perspective to the 2010 Nobel Prize. *Angew. Chem. Int. Ed.* **51**, 5062–5085 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements This research was supported by the United States National Institutes of Health, Institute of General Medical Sciences (GM-47480) and the National Science Foundation (CHE-1362763). We are grateful to A. Iimuro for experimental assistance.

Author Contributions F.M. and X.L. were involved in the discovery and development of the optimal catalysts, the corresponding methods and their various applications; S.T. and Y.S. performed the computational investigations, developed the models for the observed levels and patterns in selectivity; X.S. developed efficient routes for preparation of the dienones. A.H.H. directed the investigations and wrote the manuscript with revisions provided by the other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.H.H. (amir.hoveyda@bc.edu).

A nucleosynthetic origin for the Earth's anomalous ^{142}Nd composition

C. Burkhardt^{1,2}, L. E. Borg³, G. A. Brennecke^{2,3}, Q. R. Shollenberger^{2,3}, N. Dauphas¹ & T. Kleine²

A long-standing paradigm assumes that the chemical and isotopic compositions of many elements in the bulk silicate Earth are the same as in chondrites^{1–4}. However, the accessible Earth has a greater $^{142}\text{Nd}/^{144}\text{Nd}$ ratio than do chondrites. Because ^{142}Nd is the decay product of the now-extinct ^{146}Sm (which has a half-life of 103 million years⁵), this ^{142}Nd difference seems to require a higher-than-chondritic Sm/Nd ratio for the accessible Earth. This must have been acquired during global silicate differentiation within the first 30 million years of Solar System formation⁶ and implies the formation of a complementary ^{142}Nd -depleted reservoir that either is hidden in the deep Earth⁶, or lost to space by impact erosion^{3,7}. Whether this complementary reservoir existed, and whether or not it has been lost from Earth, is a matter of debate^{3,8,9}, and has implications for determining the bulk composition of Earth, its heat content and structure, as well as for constraining the modes and timescales of its geodynamical evolution^{3,7,9,10}. Here we show that, compared with chondrites, Earth's precursor bodies were enriched in neodymium that was produced by the slow neutron capture process (s-process) of nucleosynthesis. This s-process excess leads to higher $^{142}\text{Nd}/^{144}\text{Nd}$ ratios; after correction for this effect, the $^{142}\text{Nd}/^{144}\text{Nd}$ ratios of chondrites and the accessible Earth are indistinguishable within five parts per million. The ^{142}Nd offset between the accessible

silicate Earth and chondrites therefore reflects a higher proportion of s-process neodymium in the Earth, and not early differentiation processes. As such, our results obviate the need for hidden-reservoir or super-chondritic Earth models and imply a chondritic Sm/Nd ratio for the bulk Earth. Although chondrites formed at greater heliocentric distances and contain a different mix of presolar components than Earth, they nevertheless are suitable proxies for Earth's bulk chemical composition.

Coupled $^{146,147}\text{Sm}$ – $^{142,143}\text{Nd}$ systematics is a powerful tool to constrain the timescales and processes involved in the early differentiation of Earth, the Moon and Mars^{6,7,11–14}. However, the interpretation of ^{142}Nd signatures is complicated by the presence of nucleosynthetic isotope variations between the terrestrial planets and meteorites. Such isotope anomalies arise from the heterogeneous distribution of presolar matter at the planetary scale, and have been documented for several elements^{15–18}. Because different Nd isotopes have varying contributions from the proton process (p-process), the rapid neutron capture process (r-process), and the s-process of stellar nucleosynthesis (Extended Data Fig. 1), the observed ^{142}Nd deficits in chondrites relative to the accessible Earth could, in principle, be nucleosynthetic in origin and hence unrelated to ^{146}Sm decay^{8,16,19}. Previous studies have identified nucleosynthetic Nd (and Sm) isotope anomalies in

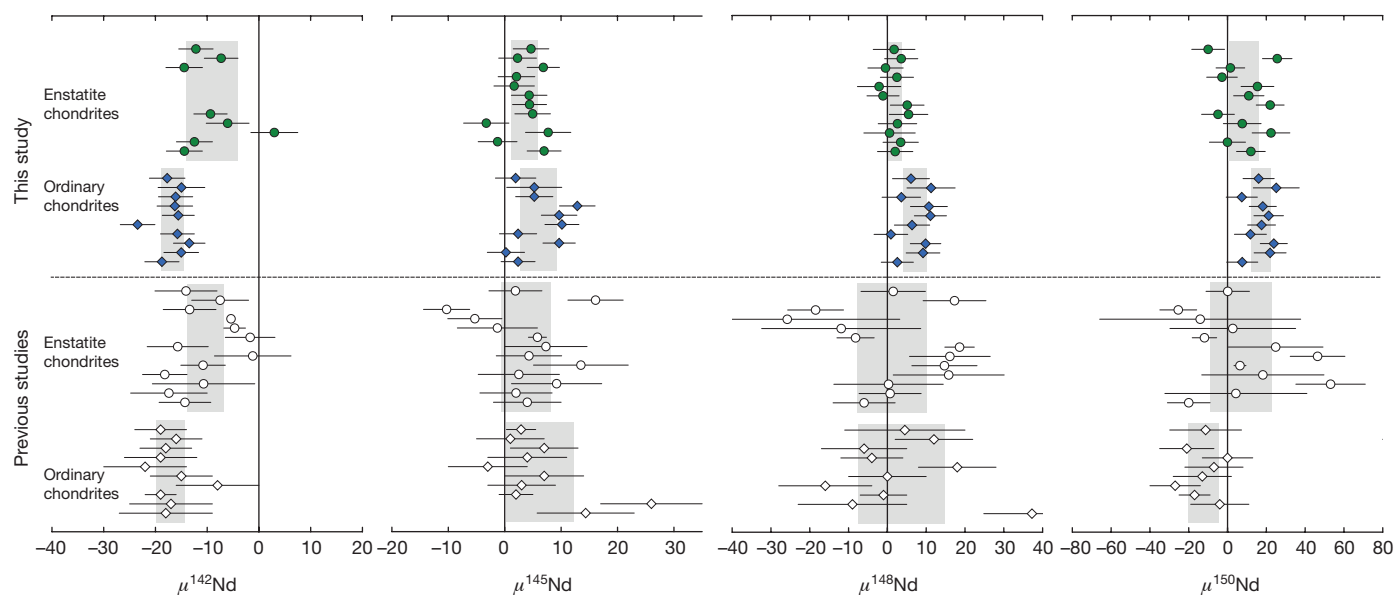


Figure 1 | Nd isotope compositions of enstatite and ordinary chondrites. Data from this study (solid symbols) show less scatter and more precisely defined mean values (the grey bars represent the 95% CI of the means from Student's *t*-values) than data from previous studies^{6,15,17,24} (open symbols), and thus reveal systematic correlated anomalies in all Nd isotopes. The uncertainties shown for the individual data points are 2 standard errors (s.e. = standard deviation/ \sqrt{n} ; where *n* is the number of cycles per

measurement) of the individual measurements. The origin of the different $\mu^{150}\text{Nd}$ values for the ordinary chondrites analysed in this study and previous studies is unclear. We note, however, that our processed standards are indistinguishable from the unprocessed JNdi-1 standard within the uncertainty, rendering it unlikely that an analytical effect in our study is responsible. Furthermore, our $\mu^{150}\text{Nd}$ data for ordinary chondrites are correlated with anomalies in other Nd isotopes, as expected.

¹Origins Laboratory, Department of the Geophysical Sciences and Enrico Fermi Institute, The University of Chicago, 5734 South Ellis Avenue, Chicago, Illinois 60637, USA. ²Institut für Planetologie, Westfälische Wilhelms-Universität Münster, Wilhelm Klemm-Strasse 10, 48149 Münster, Germany. ³Lawrence Livermore National Laboratory, L231, Livermore, California 94550, USA.

Table 1 | Sm/Nd ratios and Nd and Sm isotope compositions of meteoritic and terrestrial samples

Sample	Type	$^{147}\text{Sm}/^{144}\text{Nd}$ measured	$\mu^{142}\text{Nd}$ measured	$\mu^{142}\text{Nd}$ corrected	$\mu^{145}\text{Nd}$	$\mu^{148}\text{Nd}$	$\mu^{150}\text{Nd}$	$\mu^{144}\text{Sm}$	$\mu^{148}\text{Sm}$	$\mu^{149}\text{Sm}$	$\mu^{150}\text{Sm}$	$\mu^{154}\text{Sm}$
Hvittis (1)	EL6	0.1999(2)	−6 (5)	−12 (5)	5 (9)	2 (5)	−10 (24)	6 (22)	1 (12)	−73 (14)	159 (12)	10 (18)
Hvittis (2)	EL6	0.1986(2)	−3 (6)	−7 (6)	2 (6)	4 (7)	26 (24)	−9 (43)	−1 (10)	−76 (10)	161 (18)	2 (13)
Hvittis (3)	EL6	0.1993(2)	−10 (8)	−14 (8)	7 (13)	0 (15)	1 (31)	0 (38)	4 (10)	−75 (12)	156 (13)	−7 (11)
Atlanta (1)	EL6	0.1909(2)	−5 (6)	3 (6)	2 (6)	2 (7)	−3 (24)	14 (43)	0 (10)	−44 (10)	101 (18)	12 (13)
Atlanta (2)	EL6	0.1849(2)	−8 (8)	8 (8)	2 (13)	−2 (15)	15 (31)	−10 (43)	0 (10)	−44 (10)	91 (18)	−7 (13)
Blithfield (1)	EL6	0.2285(2)	22 (6)	−26 (6)	4 (6)	−1 (7)	11 (24)	−2 (43)	3 (10)	−34 (10)	85 (18)	1 (13)
Blithfield (2)	EL6	0.1998(2)	−9 (8)	−14 (8)	4 (13)	5 (15)	22 (31)	14 (38)	0 (10)	−51 (12)	90 (13)	3 (11)
Saint Sauveur	EH6	0.1956(2)	−10 (5)	−9 (5)	5 (9)	5 (5)	−5 (24)	−17 (22)	−6 (12)	−49 (14)	104 (12)	−4 (18)
Abee (1)	EH4	0.1874(2)	−19 (6)	−6 (6)	−3 (6)	3 (7)	8 (24)	−18 (43)	−6 (10)	−39 (10)	80 (18)	−5 (13)
Abee (2)	EH4	0.1903(2)	−5 (8)	3 (8)	8 (13)	1 (15)	22 (31)	−15 (43)	0 (10)	−33 (10)	76 (18)	−4 (13)
Indarch (1)	EH4	0.1953(2)	−14 (6)	−12 (6)	−1 (6)	3 (7)	0 (24)	18 (43)	0 (10)	−35 (10)	85 (18)	3 (13)
Indarch (2)	EH4	0.1948(2)	−16 (8)	−14 (8)	7 (13)	2 (15)	12 (31)	−7 (43)	−5 (10)	−68 (10)	130 (18)	3 (13)
Average enstatite chondrites			−10.4 (4.5)	−9.2 (4.9)	3.4 (2.1)	1.9 (1.5)	8.3 (7.4)	−2 (8)	−1 (2)			0 (4)
Kernouve	H6	0.1926(2)						18 (22)	−1 (12)	10 (14)	−3 (12)	5 (18)
Queens Mercy	H6	0.1946(2)	−20 (5)	−18 (5)	2 (9)	6 (5)	16 (24)	11 (22)	−2 (12)	11 (14)	4 (12)	12 (18)
Allegan	H5	0.1952(2)	−16 (5)	−15 (5)	5 (9)	11 (6)	25 (24)	0 (22)	−8 (12)	−12 (14)	29 (12)	3 (18)
Forest City	H5	0.1944(2)	−19 (5)	−16 (5)	5 (9)	4 (6)	7 (24)	−4 (22)	−12 (12)	0 (14)	19 (12)	11 (18)
Pultusk	H5	0.1934(2)	−20 (8)	−16 (8)	13 (13)	11 (15)	18 (31)	13 (38)	2 (10)	−52 (12)	93 (13)	−3 (11)
Sainte Marguerite (1)	H4	0.1955(2)	−16 (6)	−16 (6)	10 (6)	11 (7)	21 (24)	12 (43)	1 (10)	−16 (10)	39 (18)	1 (13)
Sainte Marguerite (2)	H4	0.1954(2)	−24 (8)	−23 (8)	10 (13)	6 (15)	18 (31)	0 (38)	−4 (10)	−18 (12)	34 (13)	−11 (11)
Bruderheim	L6	0.1935(2)	−19 (5)	−16 (5)	2 (9)	1 (5)	12 (24)	−4 (22)	2 (12)	−58 (14)	122 (12)	4 (18)
Farmington (1)	L5	0.1944(2)						26 (22)	−4 (12)	6 (14)	−1 (12)	−2 (18)
Farmington (2)	L5	0.1944(2)	−16 (6)	−13 (6)	10 (6)	10 (7)	24 (24)	−10 (43)	−2 (10)	9 (10)	6 (18)	−7 (13)
Dhurmsala	LL6	0.1965(2)	−14 (5)	−15 (5)	0 (9)	9 (5)	22 (24)	−12 (22)	5 (12)	1 (14)	23 (12)	−12 (18)
Paragould	LL5	0.1924(2)						22 (22)	−5 (12)	−70 (14)	133 (12)	2 (18)
Chelyabinsk	LL5	0.1963(2)	−18 (5)	−19 (5)	2 (9)	3 (4)	8 (24)	3 (22)	1 (12)	−8 (14)	20 (12)	4 (18)
Average ordinary chondrites			−18.3 (2.1)	−16.7 (2.0)	6.0 (3.1)	7.2 (2.7)	17.0 (4.6)	6 (7)	−2 (3)			1 (4)
Allende (1)	CV3	0.1929(2)						−85 (22)	−2 (12)	−46 (14)	97 (12)	−7 (18)
Allende (2)	CV3	0.1959(2)	−30 (5)	−30 (5)	2 (9)	9 (4)	8 (24)	−68 (22)	−8 (12)	−31 (14)	73 (12)	−2 (18)
Allende (3)	CV3	0.1961(2)	−30 (6)	−31 (6)	5 (6)	4 (7)	−6 (24)	−77 (22)	−8 (12)	−33 (14)	73 (12)	−1 (18)
Allende (4)	CV3	0.1948(2)	−33 (8)	−31 (8)	8 (13)	16 (15)	11 (31)	−89 (38)	4 (10)	−29 (12)	79 (13)	−19 (11)
Average CV			−31.3 (3.7)	−30.7 (1.1)	5.2 (7.5)	9 (16)	4 (22)	−80 (15)	−3 (9)			−7 (13)
NWA 5363	Ung.	0.2520(2)	67.1 (5.9)	−16.0 (7.5)	11 (6)	17.1 (7.3)	39 (24)	27 (43)	−1 (10)	−109 (10)	211 (18)	−4 (13)
A–ZH-5	CAI	0.2000(12)	−9.2 (7.6)	−15.2 (7.8)	−19 (13)	−28 (15)	−47 (31)	−233 (38)	62 (10)	−35 (12)	121 (13)	−43 (11)
JNdi-1 (1)	Std		0 (5)	0 (5)	−6 (9)	−5 (5)	−2 (24)					
BHVO-2	Std	0.1484(2)	−1 (5)	−1 (5)	−2 (9)	−7 (5)	−3 (24)	−7 (22)	4 (12)	6 (14)	7 (12)	−6 (18)
JNdi-1 (2)	Std		0 (8)	0 (8)	0 (13)	0 (15)	0 (31)					
BIR-1	Std	0.2759(3)	−2 (8)	−2 (8)	5 (13)	0 (15)	−10 (31)	−10 (38)	7 (10)	1 (12)	4 (13)	−17 (11)
Average of the processed standards			−0.5 (1.6)	−0.5 (1.6)	−0.7 (7.2)	−3.0 (5.6)	−3.8 (7.3)	−9 (19)	5 (12)	4 (14)	5 (18)	−11 (18)

$\mu^{142}\text{Nd} = [({}^{142}\text{Nd}/{}^{144}\text{Nd})_{\text{sample}}/({}^{142}\text{Nd}/{}^{144}\text{Nd})_{\text{standard}} - 1] \times 10^6$ and $\mu^{145}\text{Sm} = [({}^{145}\text{Sm}/{}^{152}\text{Sm})_{\text{sample}}/({}^{145}\text{Sm}/{}^{152}\text{Sm})_{\text{standard}} - 1] \times 10^6$, where all ratios have been corrected for mass fractionation by internal normalizations to fixed $^{146}\text{Nd}/^{144}\text{Nd}$ and $^{147}\text{Sm}/^{152}\text{Sm}$ ratios using the exponential law. ' $\mu^{142}\text{Nd}$ corrected' denotes $\mu^{142}\text{Nd}$ corrected for radiogenic ^{142}Nd variations to a common chondritic value ($^{147}\text{Sm}/^{144}\text{Nd} = 0.1960$). Individual sample data represent the average values of up to five measurement runs from the same filament (the full data set is available in Supplementary Information). Repeat samples (denoted 1–4) represent separate digestions that were processed through chemistry at different times and run on separate filaments. The uncertainties shown in parentheses are the external reproducibilities of the standard (2 s.d.) or two-sided Student's *t*-values 95% confidence intervals (for group averages with $n > 2$). The deficits in $\mu^{149}\text{Sm}$ values and excesses in $\mu^{150}\text{Sm}$ values that are present in some meteorite samples are due to thermal neutron capture reactions of ^{149}Sm during exposure to galactic cosmic rays, so group averages are not reported (Extended Data Fig. 5). CV refers to the Vigerano-like group of chondrites. Samples and group averages displayed in Figs 2 and 3 are in bold. Ung., ungrouped achondrite.

chondrites^{15,17} and their components^{20–23}, but these effects do not seem to fully account for the observed ^{142}Nd deficits in chondrites. For instance, although the ^{142}Nd composition of carbonaceous chondrites can partly be attributed to s- or p-process deficits^{15,17}, corrections for

these effects still leave an ^{142}Nd deficit of approximately 20 p.p.m. compared with the accessible silicate Earth. This would be consistent with Nd isotope data for bulk ordinary chondrites, which exhibit a deficit of a similar magnitude, but do not seem to show resolvable

nucleosynthetic Nd isotope anomalies^{15,17,24}. Enstatite chondrites have ^{142}Nd deficits of approximately 10 p.p.m. and also do not show clearly resolved nucleosynthetic Nd isotope anomalies²⁴. Thus, previous studies concluded that the ^{142}Nd difference between chondrites and the accessible Earth largely reflects ^{146}Sm decay and early Sm/Nd fractionation in the silicate Earth^{15,17,24}. However, this interpretation remains uncertain because the available bulk chondrite data are of insufficient precision to detect the collateral effects of nucleosynthetic heterogeneities on non-radiogenic Nd isotopes and therefore do not permit the reliable quantification of nucleosynthetic ^{142}Nd variations (Fig. 1).

Here we use high-precision Nd and Sm isotope measurements to better quantify the nucleosynthetic Nd isotope variations between chondrites and the Earth, with the ultimate goal of determining the magnitude of any radiogenic ^{142}Nd difference. We digested larger sample sizes (around 2 g) than in most previous studies, allowing us to obtain higher-precision Nd and Sm isotope data for a comprehensive set of meteorites that includes 18 chondrites, the ungrouped brachinite-like achondrite NWA 5363 and the Ca–Al-rich inclusion (CAI) A-ZH-5 from the Allende chondrite (Table 1). To evaluate the accuracy of our data we also processed the JNdi-1 standard and the terrestrial basalts BHVO-2 and BIR-1 using the same analytical procedures. Within uncertainty, the Nd and Sm isotope compositions of the processed and unprocessed standards (JNdi-1, AMES) are indistinguishable (Table 1; Figs 2, 3).

Most of the chondrites investigated cluster tightly around a ^{147}Sm – ^{143}Nd isochron at 4.568 billion years ago (Ga) (Extended Data Fig. 2a). Only the EL6 chondrites Atlanta and Blithfield plot away from the isochron, probably reflecting disturbance by late-stage impact events²⁵; the ^{142}Nd data of these samples are therefore excluded from the following discussion. After correction of measured $\mu^{142}\text{Nd}$ (for definition of $\mu^i\text{Nd}$ and $\mu^i\text{Sm}$ see Table 1) values for ^{146}Sm decay to the average chondritic $^{147}\text{Sm}/^{144}\text{Nd} = 0.1960$ (ref. 1; Extended Data Table 1), the $\mu^{142}\text{Nd}$ values are tightly clustered for each chondrite group: the enstatite chondrites define a mean $\mu^{142}\text{Nd} = -9 \pm 5$ (95% confidence interval; CI), the ordinary chondrites a mean $\mu^{142}\text{Nd} = -17 \pm 2$ (95% CI) and the Allende CV3 chondrite a mean $\mu^{142}\text{Nd} = -31 \pm 1$ (95% CI). NWA 5363 exhibits a decay-corrected $\mu^{142}\text{Nd}$ of -16 ± 7 , similar to ordinary chondrites, whereas CAI A-ZH-5 has a decay-corrected $\mu^{142}\text{Nd}$ of -15 ± 8 , consistent with the data for other Allende CAIs²².

In addition to variations in $\mu^{142}\text{Nd}$, we find resolved systematic variations in non-radiogenic Sm and Nd isotopes (Table 1, Figs 1–3). Compared with previous studies we observe less scatter for each chondrite group, reflecting the long duration and high beam intensity of our measurements, which result in more precisely defined average values for each group (Fig. 1). Plots of $\mu^{145}\text{Nd}$ and $\mu^{150}\text{Nd}$ versus $\mu^{148}\text{Nd}$ reveal positively correlated anomalies, with the enstatite chondrites being closest to the terrestrial value, followed by carbonaceous and ordinary chondrites, and then NWA 5363 (Fig. 2a, b). The meteorite samples plot along mixing lines between terrestrial Nd (that is, $\mu^i\text{Nd} = 0$) and pure s-process Nd, regardless of whether the s-process composition is derived from presolar SiC grains²⁶, nucleosynthesis models²⁷ or data for acid leachates of primitive chondrites^{20,21}. Thus, the variability in non-radiogenic Nd isotopes among the meteorites reflects variable s-deficits relative to the Earth, consistent with inferences from other elements^{16,28,29}.

The $\mu^{145}\text{Nd}$, $\mu^{148}\text{Nd}$ and $\mu^{150}\text{Nd}$ anomalies of Allende are similar to those of ordinary and enstatite chondrites, although for most other elements nucleosynthetic anomalies are typically largest in carbonaceous chondrites^{16,18,28–30}. The reason for the subdued Nd isotopic anomalies in Allende is the presence of CAIs, which host about half of the Nd and Sm in Allende³¹, and which, for these elements, are characterized by an s-excess and a p-deficit (Figs 2, 3). Mass balance calculations (see Methods and Extended Data Table 2) indicate that the composition of a CAI-free Allende would have $\mu^{145}\text{Nd}$, $\mu^{148}\text{Nd}$ and $\mu^{150}\text{Nd}$ values of 27 ± 14 , 39 ± 28 , and 56 ± 41 , respectively; these anomalies are larger than those of ordinary and enstatite chondrites and thus imply that before addition of CAIs, the Allende chondrite

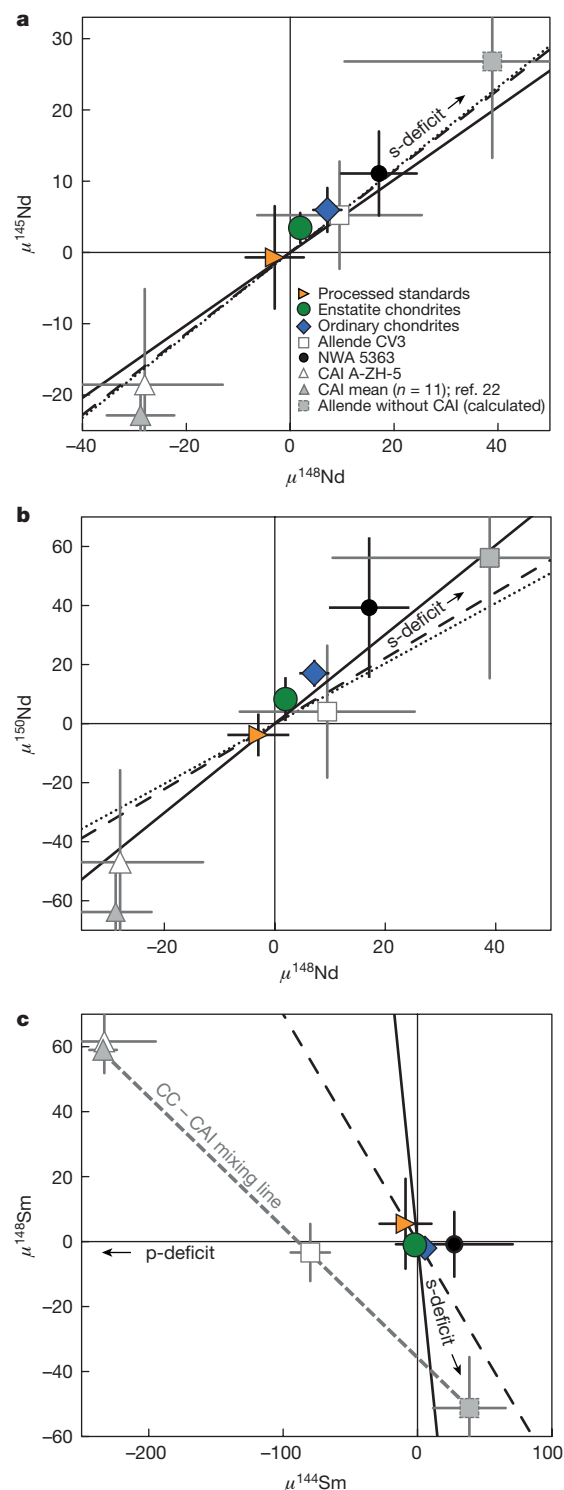


Figure 2 | Nd and Sm isotope variations among meteoritic and terrestrial samples. **a**, Anomalies in non-radiogenic Nd isotopes $\mu^{145}\text{Nd}$ and $\mu^{148}\text{Nd}$ are consistent with a heterogeneous distribution of s-process Nd. Solid, dotted and dashed lines are the mixing lines between terrestrial Nd and s-process Nd, calculated using modelled s-process compositions²⁷, Nd data for presolar SiC grains²⁶ and Nd data for chondrite leachates^{20,21}, respectively. The isotopic composition measured for bulk Allende can be accounted for by the admixture of CAIs in a CAI-free carbonaceous chondrite (CC) source reservoir (the 'Allende without CAI (calculated)' data point) that is characterized by an s-process deficit. **b**, Same as **a** but for $\mu^{148}\text{Nd}$ and $\mu^{150}\text{Nd}$. **c**, The p-deficit observed for bulk Allende in $\mu^{144}\text{Sm}$ can also be attributed to the admixture of CAIs. The grey dashed CC – CAI line represents a mixing line calculated by subtracting CAIs from the isotopic composition measured for bulk Allende. Error bars indicate the 95% CI.

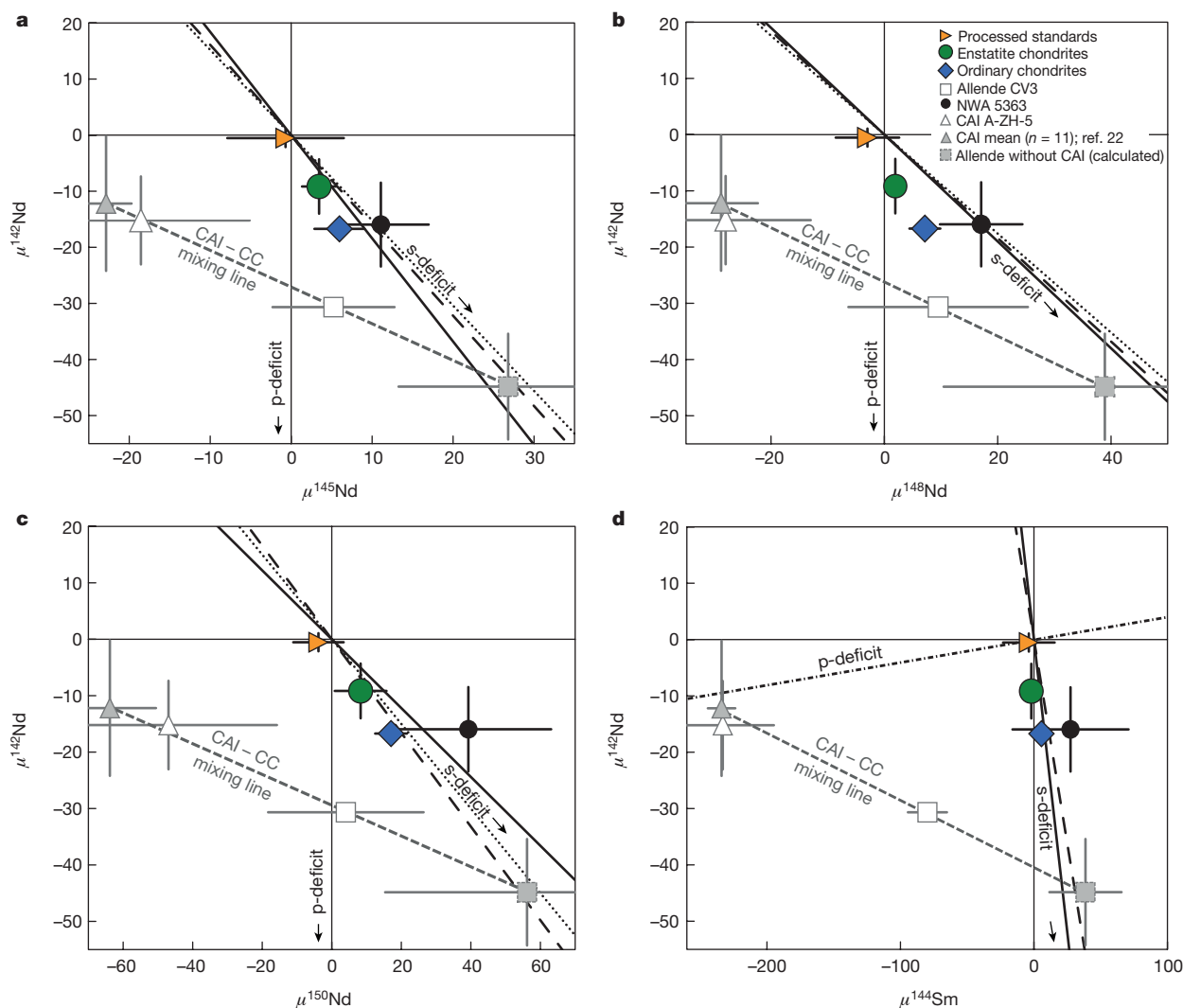


Figure 3 | Nd and Sm isotope variations among meteoritic and terrestrial samples. **a**, For enstatite chondrites, ordinary chondrites and NWA 5363, the $\mu^{142}\text{Nd}$ anomalies are correlated with the non-radiogenic $\mu^{145}\text{Nd}$ anomalies, as expected for a heterogeneous distribution of s-process Nd. The Allende carbonaceous chondrite plots off this

had a substantial s-deficit (Fig. 2a, b). This interpretation is consistent with Sm isotope data for Allende and other carbonaceous chondrites, because the calculated CAI-free composition of these chondrites also shows an s-deficit (Fig. 2c, Extended Data Fig. 3). Thus, the displacement of the carbonaceous chondrites from the s-deficit line that is defined by ordinary and enstatite chondrites reflects the admixture of CAIs to carbonaceous chondrites. Note that, for ordinary and enstatite chondrites the effects of admixing CAIs are estimated to be no larger than 2 p.p.m. for Nd isotope ratios and 5 p.p.m. for Sm isotope ratios (Methods, Extended Data Table 2), and that the expected s-process Sm isotope anomalies ($<10\mu^{144}\text{Sm}$ and $>-20\mu^{148}\text{Sm}$) for these two groups of chondrites are too small to be resolvable with the analytical precision of our Sm isotope measurements.

Using the information gained from the non-radiogenic isotopes, we can now assess the effect of nucleosynthetic anomalies on $\mu^{142}\text{Nd}$. The bulk meteorite data show inverse correlations between $\mu^{142}\text{Nd}$ and $\mu^{145}\text{Nd}$, $\mu^{148}\text{Nd}$, $\mu^{150}\text{Nd}$ and $\mu^{144}\text{Sm}$ (Fig. 3), which are consistent with the covariations expected from a heterogeneous distribution of s-process isotopes. Enstatite and ordinary chondrites, as well as NWA 5363, plot on mixing lines between terrestrial and s-process Nd. The Allende CV3 chondrite is displaced from these correlations owing to the admixture of CAIs, and a calculated CAI-free carbonaceous chondrite composition plots on the s-mixing line defined by the other meteorites (Fig. 3).

correlation owing to the admixture of CAIs. Mass balance calculations indicate that a CAI-free carbonaceous chondrite source reservoir was characterized by an s-process deficit. **b**, Same as **a** but in $\mu^{142}\text{Nd}$ versus $\mu^{148}\text{Nd}$ space. **c**, Same as **a** but in $\mu^{142}\text{Nd}$ versus $\mu^{150}\text{Nd}$ space. **d**, Same as **a** but in $\mu^{142}\text{Nd}$ versus $\mu^{144}\text{Sm}$ space. Error bars indicate the 95% CI.

The slopes obtained from linear regressions of the bulk meteorites (excluding Allende) are in good agreement with those calculated for mixing lines between terrestrial and s-process Nd, regardless of which estimate for the s-process composition is used^{20,21,26,27} and whether or not the calculated CAI-free carbonaceous chondrite composition and the processed standards are included in the regressions (Extended Data Fig. 4). The intercept values obtained from the regressions can thus be used to determine $\mu^{142}\text{Nd}$ values corrected for s-process heterogeneity. For all regressions the intercept values are indistinguishable from each other, with an average value of approximately -5 p.p.m. relative to the JNdi-1 standard (Extended Data Table 3). Alternatively, $\mu^{142}\text{Nd}$ values corrected for nucleosynthetic anomalies can be calculated for each meteorite group separately, using their measured $\mu^{145}\text{Nd}$, $\mu^{148}\text{Nd}$ and $\mu^{150}\text{Nd}$ values combined with the slopes of the s-mixing lines. Regardless of which s-process mixing relationships are applied, the calculated $\mu^{142}\text{Nd}_{\text{s-corrected}}$ values are all mutually consistent and indistinguishable from each other (Extended Data Table 3), resulting in an average $\mu^{142}\text{Nd}_{\text{s-corrected}} = -5 \pm 2$ p.p.m. Although this value is slightly negative, it is within the approximate long-term ± 5 p.p.m. reproducibility of the JNdi-1 standard. When the regressions and corrections are calculated relative to the mean Nd isotope composition measured for the processed terrestrial standards, $\mu^{142}\text{Nd}_{\text{s-corrected}}$ reduces to -2 ± 2 p.p.m. (Extended Data Table 3). We conclude that

after correction for nucleosynthetic Nd isotope heterogeneity, the ^{142}Nd compositions of chondrites and the accessible silicate Earth are indistinguishable at the current level of analytical precision (approximately 5 p.p.m.).

The lack of a resolved radiogenic ^{142}Nd difference between chondrites and the accessible silicate Earth supports the long-standing paradigm of a chondritic Sm/Nd for the bulk Earth and requires the revision of the conclusions from several previous studies about the early differentiation, composition, structure and heat budget of the Earth. These studies interpreted the ^{142}Nd offset between chondrites and terrestrial samples as resulting from ^{146}Sm -decay and an early global Sm/Nd fractionation in the Earth's mantle^{3,6,7,9,10}. Our results instead demonstrate that chondrites and the accessible Earth have indistinguishable radiogenic ^{142}Nd compositions, refuting the evidence for an early global silicate differentiation of the Earth and indicating that the hidden, enriched reservoir hypothesized in earlier studies^{3,6,9,10} does not exist. Moreover, our results rule out the extensive loss of early-formed crust by collisional erosion^{3,7,9}, because otherwise the bulk silicate Earth would not have a chondritic Sm/Nd ratio. Finally, the evidence for chondritic Sm/Nd ratios in the bulk Earth implies chondritic abundances of other refractory elements, including the heat-producing elements U and Th. Thus, the total radiogenic heat generated over Earth's history is almost a factor of two higher than was estimated recently for a non-chondritic composition of the Earth⁹.

Our results demonstrate that chondrites are the most appropriate proxy for the elemental composition of the Earth. However, they also highlight that chondrites cannot be the actual building blocks of the Earth because they are deficient in a presolar component that contains s-process matter. The s-process deficit increases from enstatite via ordinary to carbonaceous chondrites, indicating that the distribution of presolar matter in the solar protoplanetary disk either varied as a function of heliocentric distance or over time. For instance, the nucleosynthetic isotope heterogeneity within the disk may reflect differences in the thermal processing of stellar-derived dust, imparting isotopic heterogeneity on an initially homogeneous disk, but it could also reflect distinct compositions of material added to the disk from the molecular cloud at different times^{18,28–30}. Either way, the increasing deficit in s-process matter with increasing heliocentric distance provides a new means for identifying genetic relationships among planetary bodies. For instance, Mars formed at a greater heliocentric distance than Earth and should, therefore, be characterized by an s-process deficit that may be similar to those observed for enstatite and ordinary chondrites. Thus, high-precision Nd isotopic data for Martian meteorites will make it possible to determine the distinct sources of the building materials of Earth and Mars. This information is not only critical for dating the differentiation of Mars¹³, but also for testing models of terrestrial planet formation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 January; accepted 3 June 2016.

1. Bouvier, A., Vervoort, J. D. & Patchett, P. J. The Lu–Hf and Sm–Nd isotopic composition of CHUR: constraints from unequilibrated chondrites and implications for the bulk composition of terrestrial planets. *Earth Planet. Sci. Lett.* **273**, 48–57 (2008).
2. Nakamura, N. Determination of REE, Ba, Fe, Mg, Na, and K in carbonaceous and ordinary chondrites. *Geochim. Cosmochim. Acta* **38**, 757–775 (1974).
3. Campbell, I. H. & O'Neill, H. S. C. Evidence against a chondritic Earth. *Nature* **483**, 553–558 (2012).
4. Jacobsen, S. B. & Wasserburg, G. J. Sm–Nd isotopic evolution of chondrites. *Earth Planet. Sci. Lett.* **50**, 139–155 (1980).
5. Meissner, F., Schmidt-Ott, W.-D. & Ziegler, L. Half-life and α -ray energy of ^{146}Sm . *Z. Phys. A* **327**, 171–174 (1987).
6. Boyet, M. & Carlson, R. W. ^{142}Nd evidence for early (>4.53 Ga) global differentiation of the silicate Earth. *Science* **309**, 576–581 (2005).
7. Caro, G., Bourdon, B., Halliday, A. & Quitté, G. Superchondritic Sm/Nd in Mars, Earth and the Moon. *Nature* **452**, 336–339 (2008).

8. Huang, S., Jacobsen, S. B. & Mukhopadhyay, S. ^{147}Sm – ^{143}Nd systematics of Earth are inconsistent with a superchondritic Sm/Nd ratio. *Proc. Natl Acad. Sci. USA* **110**, 4929–4934 (2013).
9. Jellinek, A. M. & Jackson, M. G. Connections between the bulk composition, geodynamics and habitability of Earth. *Nat. Geosci.* **8**, 587–593 (2015).
10. Carlson, R. W. & Boyet, M. Composition of the Earth's interior: the importance of early events. *Phil. Trans. R. Soc. A* **366**, 4077–4103 (2008).
11. Bennett, V. C., Brandon, A. D. & Nutman, A. P. Coupled ^{142}Nd – ^{143}Nd isotopic evidence for Hadean mantle dynamics. *Science* **318**, 1907–1910 (2007).
12. Brandon, A. D. *et al.* Re-evaluating ^{142}Nd / ^{144}Nd in lunar mare basalts with implications for the early evolution and bulk Sm/Nd of the Moon. *Geochim. Cosmochim. Acta* **73**, 6421–6445 (2009).
13. Debaille, V., Brandon, A. D., Yin, Q. Z. & Jacobsen, B. Coupled ^{142}Nd – ^{143}Nd evidence for a protracted magma ocean in Mars. *Nature* **450**, 525–528 (2007).
14. Harper, C. L. & Jacobsen, S. B. Evidence from coupled ^{147}Sm – ^{143}Nd and ^{146}Sm – ^{142}Nd systematics for very early (4.5 Gyr) differentiation of the Earth's mantle. *Nature* **360**, 728–732 (1992).
15. Andreasen, R. & Sharma, M. Solar nebula heterogeneity in p-process samarium and neodymium isotopes. *Science* **314**, 806–809 (2006).
16. Burkhardt, C. *et al.* Molybdenum isotope anomalies in meteorites: Constraints on solar nebula evolution and origin of the Earth. *Earth Planet. Sci. Lett.* **312**, 390–400 (2011).
17. Carlson, R. W., Boyet, M. & Horan, M. F. Chondrite barium, neodymium, and samarium isotopic heterogeneity and early Earth differentiation. *Science* **316**, 1175–1178 (2007).
18. Trinquier, A. *et al.* Origin of nucleosynthetic isotope heterogeneity in the solar protoplanetary disk. *Science* **324**, 374–376 (2009).
19. Sprung, P., Kleine, T. & Scherer, E. E. Isotopic evidence for chondritic Lu/Hf and Sm/Nd of the Moon. *Earth Planet. Sci. Lett.* **380**, 77–87 (2013).
20. Boyet, M. & Gannoun, A. Nucleosynthetic Nd isotope anomalies in primitive enstatite chondrites. *Geochim. Cosmochim. Acta* **121**, 652–666 (2013).
21. Qin, L. P., Carlson, R. W. & Alexander, C. M. O. Correlated nucleosynthetic isotopic variability in Cr, Sr, Ba, Sm, Nd and Hf in Murchison and QUE 97008. *Geochim. Cosmochim. Acta* **75**, 7806–7828 (2011).
22. Brennecka, G. A., Borg, L. E. & Wadhwa, M. Evidence for supernova injection into the solar nebula and the decoupling of r-process nucleosynthesis. *Proc. Natl Acad. Sci. USA* **110**, 17241–17246 (2013).
23. Marks, N. E., Borg, L. E., Hutcheon, I. D., Jacobsen, B. & Clayton, R. N. Samarium–neodymium chronology and rubidium–strontium systematics of an Allende calcium–aluminum-rich inclusion with implications for ^{146}Sm half-life. *Earth Planet. Sci. Lett.* **405**, 15–24 (2014).
24. Gannoun, A., Boyet, M., Rizo, H. & El Goresy, A. ^{146}Sm – ^{142}Nd systematics measured in enstatite chondrites reveals a heterogeneous distribution of ^{142}Nd in the solar nebula. *Proc. Natl Acad. Sci. USA* **108**, 7693–7697 (2011).
25. Rubin, A. E. Impact features of enstatite-rich meteorites. *Chem. Erde-Geochem.* **75**, 1–28, (2015).
26. Hoppe, P. & Ott, U. Mainstream silicon carbide grains from meteorites. *AIP Conf. Proc.* **402**, 27–58 (1997).
27. Arlandini, C., Käppeler, F. & Wisshak, K. Neutron capture in low-mass asymptotic giant branch stars: cross sections and abundance signatures. *Astrophys. J.* **525**, 886–900 (1999).
28. Akram, W., Schönbächler, M., Bisterzo, S. & Gallino, R. Zirconium isotope evidence for the heterogeneous distribution of s-process materials in the solar system. *Geochim. Cosmochim. Acta* **165**, 484–500 (2015).
29. Fischer-Gödde, M., Burkhardt, C., Kruijer, T. S. & Kleine, T. Ru isotope heterogeneity in the solar protoplanetary disk. *Geochim. Cosmochim. Acta* **168**, 151–171 (2015).
30. Dauphas, N. *et al.* Calcium-48 isotopic anomalies in bulk chondrites and achondrites: evidence for a uniform isotopic reservoir in the inner protoplanetary disk. *Earth Planet. Sci. Lett.* **407**, 96–108 (2014).
31. Stracke, A. *et al.* Refractory element fractionation in the Allende meteorite: implications for solar nebula condensation and the chondritic composition of planetary bodies. *Geochim. Cosmochim. Acta* **85**, 114–141 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the Field Museum for providing samples, S.-G. Lee for help setting up the chemistry in Chicago, R. Carlson for discussions. This work was funded through SNF PBE2PZ-145946 (CB); NASA (NNX14AK09G, OJ-30381-0036A, NNX15AJ25G), NSF (EAR144495, EAR150259) (ND); NASA NNH12AT84I (LB) and the ERC (Grant Agreement 616564 'ISOCORE') (TK). The work performed by L.E.B., G.A.B. and Q.R.S. was done under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Author Contributions C.B. initiated the project in collaboration with L.E.B., N.D. and T.K., acquired and processed the samples in Chicago and wrote a first draft of the manuscript. L.E.B., G.A.B. and Q.R.S. performed additional chemistry and measured all samples in Livermore. All authors contributed to the data interpretation and editing of the manuscript.

Author Information Data are available at the EarthChem library (<http://dx.doi.org/10.1594/IEDA/100597>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.B. (burkhardt@uni-muenster.de).

METHODS

Samples. To avoid the artefacts that can be associated with incomplete dissolution of refractory presolar components and to minimize potential disturbances through terrestrial alteration, only equilibrated chondrites (petrologic classes 4–6; except the CV3 Allende) from observed falls were selected for this study. Equilibrated chondrites are devoid of presolar grains because these components were destroyed during thermal metamorphism on the meteorite parent body³²; for Allende (metamorphic grade of 3.2 to >3.6), which may contain trace amounts of presolar grains³², no difference in Nd isotopic composition was observed between table-top acid-digested, bomb-digested and alkali-fused samples^{6,17}, indicating that for this meteorite all Nd carriers are accessed by standard acid digestion. Our sample set includes eleven ordinary chondrites (six H, two L and three LL), six enstatite chondrites (three EL and three EH), the carbonaceous chondrite Allende and the brachinite-like achondrite NWA 5363, which is a melt-depleted ultramafic sample from a partially differentiated asteroid³³. This brachinite-like sample was added to the study because of its unique isotope anomalies: while the oxygen and nickel isotopic compositions of NWA 5363 are indistinguishable from terrestrial values, it exhibits nucleosynthetic isotope anomalies in Ti, Ca, Mo and Ru that are closer to ordinary chondrites³⁴. In addition to bulk meteorites, we analysed the CAI A-ZH-5 from the Allende chondrite and, to evaluate the accuracy of our analytical methods, we also processed the JNdi-1 standard, as well as the terrestrial basalt standards BHVO-2 and BIR-1.

Sample preparation and chemical separation of Nd and Sm. Meteorite pieces were cleaned with abrasive paper, ultrasonicated in methanol and subsequently crushed to a fine powder in an acid-cleaned agate mortar exclusively used for meteorite work at the Origins Laboratory, Chicago. For each analysis about 2 g of meteorite powder was digested in a HF–HNO₃–HClO₄ mixture and aqua regia in 90 ml Savillex teflon vials for about 10 d on a hotplate at 170 °C. After several dry-downs, ultrasonication and redissolution steps in aqua regia and HCl, the samples were redissolved in HCl and, once a clear solution was obtained, an aliquot of approximately 5% was taken for Sm and Nd concentration measurements by isotope dilution.

Chemical procedures for Sm and Nd concentration measurements. The 5% aliquots were sent from the Origins Laboratory to the Lawrence Livermore National Laboratory (LLNL), where they were equilibrated with a ¹⁴⁹Sm–¹⁵⁰Nd mixed isotopic tracer. Rare earth elements (REE) were purified from the matrix of these aliquots using 2 ml BioRad columns filled with AG50-X8 (200–400 mesh) resin and 2 M and 6 M HCl. The REE were further purified using 150 µl Teflon columns with RE-Spec resin and 1 M and 0.05 M HNO₃. Sm and Nd were purified from the other REE using 15 cm glass columns, Ln-Spec resin and 0.25 M and 0.60 M HCl. Total blanks of the isotope dilution procedures were 25 pg of Nd and 8 pg of Sm, resulting in Nd and Sm sample-to-blank ratios greater than 1,500 for all but one sample. The blank corrections resulted in shifts in the ¹⁴⁷Sm/¹⁴⁴Nd ratios that were less than 0.003% and thus substantially smaller than the typical uncertainty of 0.1% associated with the isotope dilution measurements. For NWA 5363, the Nd and Sm sample-to-blank ratios were 751 and 760, respectively, and thus required a blank correction of 0.13% for the Nd and Sm concentrations (for example, the reported 0.112 p.p.m. Nd abundance was corrected by 0.00015 p.p.m.). The blank correction is reflected in the larger uncertainty of 0.2% on the ¹⁴⁷Sm/¹⁴⁴Nd of NWA 5363.

Chemical procedures for Sm and Nd isotope composition measurements. After aliquoting, the remaining 95% of the sample solution was reduced and HNO₃ was added. The REE cut of CAI A-ZH-5 that was obtained in a previous study³⁵ (where the digested sample was processed through an anion exchange chromatography to separate Ti, Zr, Hf, W and Mo from the matrix; for details see ref. 35) was added to the project at this point. After additional dry-downs in aqua regia and HNO₃, samples were redissolved in approximately 35 ml of 3 M HNO₃ and 350 mg of H₃BO₃ was added before the solutions were centrifuged. A fine-grained, black, low-density residue (probably carbon-based) was present for some of the chondrites at this point and was discarded; note that because we analysed equilibrated chondrites, this C-bearing phase does not contain presolar material and therefore does not influence the Nd isotopic composition of the non-radiogenic isotopes. Furthermore, changes in the Sm/Nd ratios or the radiogenic Nd isotopic signatures of the samples by this material is also excluded, given the very good agreement of our decay-corrected ¹⁴²Nd and ¹⁴³Nd data with previous studies (Fig. 1; Extended Data Fig. 2). After centrifugation, the REE were separated from the matrix elements by loading the solutions onto two 2 ml Eichrom TODGA ion exchange columns stacked on top of each other³⁶. To further purify the REE cut, the separation was repeated using a single 2 ml TODGA column. Separation of Sm and Nd from interfering REEs was accomplished with 0.2 cm × 25 cm quartz columns with AG50W-X8 (NH₄⁺ form, with a pH of around 7) as the stationary phase and 0.2 M alpha-hydroxyisobutyric acid (pH adjusted to 4.6) as the fluid phase. The Sm and Nd cuts were passed twice over this column at the University of Chicago and were then sent to LLNL. The Nd was further purified at LLNL using 0.2 M alpha-hydroxyisobutyric acid adjusted to a pH of 4.40 on pressurized quartz glass columns loaded with AG50W-X8 (NH₄⁺ form) resin. Neodymium

was separated from the alpha-hydroxyisobutyric acid using 2 ml columns loaded with AG50W-X8 (200–400 mesh) resin using water, 2 M HCl, and 6 M HCl. The yields of the chemical procedure were determined by ICP-MS on small aliquots of the processed Nd and Sm cuts and ranged between 62% and 95% for Nd (with a mean yield of 80%) and 56% and 98% for Sm (with a mean yield of 75%). The variable yields do not have any noticeable influence on the measured Nd and Sm isotopic compositions. This is indicated both by the fact that several samples processed multiple times displayed variable yields, but had very homogeneous isotopic compositions, and that the terrestrial rock standards passed through the chemistry have indistinguishable compositions from the unprocessed standard. These observations further suggest that either the exponential law is well-suited to correct any yield-related induced mass-dependent isotope variations, or that the sample loss is associated with processes that do not induce mass-dependent fractionation effects, for example, pipetting of the samples on the columns or the loss of dry sample material from the beakers by static effects. The latter erratic losses seem to be the most likely explanation for the variable yields, which vary in a non-systematic way within a chemical campaign and among multiple digestions of the same meteorites. The procedural blanks associated with Nd and Sm isotope composition measurements were 50 pg and 12 pg respectively, and thus contributed negligibly (<0.03% of total analyte) to the isotope compositions of the samples, requiring no corrections to be made.

Procedures of Nd and Sm isotope measurement by TIMS. The Nd isotope compositions were analysed using a ThermoScientific Triton thermal ionization mass spectrometer (TIMS) at LLNL. Neodymium was loaded on zone-refined Re filaments in 2 M HCl and analysed as Nd⁺ using a second Re ionization filament. Isotope ratios were measured using a two-mass-step procedure that calculates ¹⁴²Nd/¹⁴⁴Nd and ¹⁴⁸Nd/¹⁴⁴Nd ratios dynamically, while measuring the other Nd isotopes statically following a modified version of previously established procedures¹⁷. The cup configuration of lines 1 and 2 are: L3 = ¹⁴²Nd, L2 = ¹⁴³Nd, L1 = ¹⁴⁴Nd, C = ¹⁴⁵Nd, H1 = ¹⁴⁶Nd, H2 = ¹⁴⁸Nd, H3 = ¹⁴⁹Sm and H4 = ¹⁵⁰Nd, and L3 = ¹⁴⁰Ce, L2 = ¹⁴¹Pr, L1 = ¹⁴²Nd, C = ¹⁴³Nd, H1 = ¹⁴⁴Nd, H2 = ¹⁴⁶Nd, H3 = ¹⁴⁷Sm and H4 = ¹⁴⁸Nd, respectively. Individual mass spectrometer runs consisted of 540 ratios of 8 s integrations. The dynamic ¹⁴²Nd/¹⁴⁴Nd ratio is calculated from ¹⁴²Nd/¹⁴⁴Nd measured in cycle 2 normalized to ¹⁴⁶Nd/¹⁴⁴Nd measured in cycle 1, whereas the dynamic ¹⁴⁸Nd/¹⁴⁴Nd ratio is calculated from the ¹⁴⁸Nd/¹⁴⁶Nd ratio measured in cycle 1 normalized to ¹⁴⁶Nd/¹⁴⁴Nd measured in cycle 2. The ¹⁴³Nd/¹⁴⁴Nd ratio is calculated from the average of the 1,080 ratios of data collected in cycles 1 and 2. The ¹⁴⁵Nd/¹⁴⁴Nd ratio represents the average of 540 ratios collected in cycle 1. Most samples were run at least twice from the same filaments. Signal sizes varied from ¹⁴⁴Nd = 3.2 × 10^{–11} A to 5.4 × 10^{–11} A, with most averaging in excess of 4.3 × 10^{–11} A. Fractionation was corrected by assuming ¹⁴⁶Nd/¹⁴⁴Nd = 0.7219 using the exponential law. The Nd isotope data were acquired in three measurement campaigns that were separated by a cup exchange and maintenance work on the Triton. To avoid any bias that might have been introduced by these events, the data obtained in each of the campaigns were normalized to the mean JNdi-1 composition measured in the respective campaign (see Supplementary Information). The external reproducibility of the standard (2 s.d.) for ¹⁴²Nd/¹⁴⁴Nd, ¹⁴⁵Nd/¹⁴⁴Nd, ¹⁴⁸Nd/¹⁴⁴Nd and ¹⁵⁰Nd/¹⁴⁴Nd are: 5 p.p.m., 9 p.p.m., 3 p.p.m. and 24 p.p.m. in campaign 1; 6 p.p.m., 6 p.p.m., 7 p.p.m. and 24 p.p.m. in campaign 2; and 8 p.p.m., 13 p.p.m., 15 p.p.m. and 31 p.p.m. in campaign 3. Table 1 presents the average values of multiple measurements from the same filament. The associated uncertainties represent the external reproducibility (2 s.d.) of the standard during that campaign, or the uncertainty of the sample measurements (2 s.e.), which were larger than the external reproducibility of the standard (3 p.p.m.) for some of the ¹⁴⁸Nd/¹⁴⁴Nd sample runs in campaign 1. Interferences from Ce and Sm are monitored at ¹⁴⁰Ce and ¹⁴⁹Sm and are presented in Supplementary Table 1.

Samarium was loaded in 2 M HCl onto a zone-refined Re filament and analysed as Sm⁺ using double Re filaments. All Sm isotopes, along with the interferences from Nd (measured as ¹⁴⁶Nd) were measured statically for 200 ratios of 8 s integration each. Instrument fractionation was corrected by assuming ¹⁴⁷Sm/¹⁵²Sm = 0.56803 using the exponential law. The cup configuration for Sm isotope composition measurements is: L4 = ¹⁴⁴Sm, L3 = ¹⁴⁶Nd, L2 = ¹⁴⁷Sm, L1 = ¹⁴⁸Sm, C = ¹⁴⁹Sm, H1 = ¹⁵⁰Sm, H2 = ¹⁵²Sm, H3 = ¹⁵⁴Sm and H4 = ¹⁵⁵Gd. Sample measurements consisted of one to three static runs from the same filament, depending on the amount of Sm available, and were obtained at (1–2) × 10^{–11} A ¹⁴⁹Sm. The data were acquired in three campaigns and are given in the Supplementary Information. Samarium isotope anomalies were calculated relative to the mean composition of the AMES Sm standard analysed in each campaign (see Supplementary Information). The external reproducibility of the standard for ¹⁴⁴Sm/¹⁵²Sm, ¹⁴⁸Sm/¹⁵²Sm, ¹⁴⁹Sm/¹⁵²Sm, ¹⁵⁰Sm/¹⁵²Sm and ¹⁵⁴Sm/¹⁵²Sm are: 22 p.p.m., 12 p.p.m., 14 p.p.m., 12 p.p.m. and 18 p.p.m. in campaign 1; 43 p.p.m., 10 p.p.m., 10 p.p.m., 18 p.p.m. and 13 p.p.m. in campaign 2; and 38 p.p.m., 10 p.p.m., 12 p.p.m., 13 p.p.m. and 11 p.p.m. in campaign 3. Table 1 presents average

values of the multiple measurements run from the same filament, and the reported uncertainties are 2 s.d. of the standard.

The Nd and Sm concentrations were determined using a ThermoScientific TIMS in static mode. Measurements consisted of 200 cycles with 8 s integration time each. Concentration data and $^{147}\text{Sm}/^{144}\text{Nd}$ ratios are given in Supplementary Table 3. Note that the nucleosynthetic anomalies measured here have no noticeable effect on the accuracy and precision of the Sm and Nd concentration measurement (the minimum variation in the Sm and Nd isotopic compositions that would be required to shift the $^{147}\text{Sm}/^{144}\text{Nd}$ ratios beyond uncertainty are 270 μSm and 560 μNd units, respectively; and thus substantially larger than the deviations we observed).

Isotopic mass balance calculations between CAIs and Allende. CAIs found in carbonaceous chondrites are considered to be the oldest surviving objects to have formed in the solar nebula, presumably by condensation from nebular gas. They often exhibit isotopic anomalies that are substantially different from their chondrite host rocks^{16,18,22,23,37}, strongly suggesting that they are not genetically related to the reservoir from which the other chondrite components (namely chondrules and matrix) originated. The Nd and Sm isotopic composition of bulk carbonaceous chondrites is thus most likely to be influenced by CAIs, especially since the (light) REEs in these objects are enriched relative to the host rocks (for example, up to around 20 \times for CAIs from the CV (Vigarano-like) chondrite group, up to around 100 \times for CAIs from the CM (Mighei-like) chondrite group).

Indeed, our measurements imply that CAI material exert a strong control on the Nd and Sm isotope composition of bulk carbonaceous chondrites, because our Allende data (as well as literature data of carbonaceous chondrites) are displaced towards the CAI composition in μNd versus μNd , μNd versus μSm and μSm versus μSm (with i and j representing different isotopes of the Nd and Sm diagrams (Figs 2, 3; Extended Data Fig. 3)).

To quantify the effect of CAIs on the Allende composition and characterize the composition of the CAI-free carbonaceous chondrite source reservoir we performed an isotopic mass balance calculation. For Nd this has the form

$$\text{Nd}_{\text{Allende}} = X\text{Nd}_{\text{source}} + (1 - X)\text{Nd}_{\text{CAI}} \quad (1)$$

where $\text{Nd}_{\text{Allende}}$ is the concentration of Nd in Allende, which is given by the sum of Nd in the carbonaceous chondrite source reservoir ($\text{Nd}_{\text{source}}$) and the Nd contributed by the CAIs (Nd_{CAI}) and X is the fraction of non-CAI material in Allende.

For the isotopic composition we can likewise write

$$\mu^i\text{Nd}_{\text{Allende}}\text{Nd}_{\text{Allende}} = X\mu^i\text{Nd}_{\text{source}}\text{Nd}_{\text{source}} + (1 - X)\mu^i\text{Nd}_{\text{CAI}}\text{Nd}_{\text{CAI}} \quad (2)$$

Using the isotopic compositions measured for Allende (this study), Allende CAIs (the mean value of 11 CAIs reported in ref. 22) and 3 wt% CAIs in Allende³⁸, and mean Nd concentrations of 0.967 p.p.m. and 14 p.p.m. for Allende and Allende CAIs³¹, we can solve for the unknown concentration and isotopic composition of the CAI-free material according to

$$\text{Nd}_{\text{source}} = \frac{\text{Nd}_{\text{Allende}} - (1 - X)\text{Nd}_{\text{CAI}}}{X} \quad (3)$$

and

$$\mu^i\text{Nd}_{\text{source}} = \frac{\mu^i\text{Nd}_{\text{Allende}}\text{Nd}_{\text{Allende}} - (1 - X)\mu^i\text{Nd}_{\text{CAI}}\text{Nd}_{\text{CAI}}}{\text{Nd}_{\text{Allende}} - (1 - X)\text{Nd}_{\text{CAI}}} \quad (4)$$

The uncertainty in $\mu^i\text{Nd}_{\text{source}}$ is mainly determined by the uncertainties in the measured isotopic compositions of Allende and the CAIs and was calculated by propagating them according to

$$\sigma_{\mu^i\text{Nd}_{\text{source}}}^2 = \left(\frac{\partial F_{\mu^i\text{Nd}_{\text{source}}}}{\partial \mu^i\text{Nd}_{\text{Allende}}} \right)^2 \sigma_{\mu^i\text{Nd}_{\text{Allende}}}^2 + \left(\frac{\partial F_{\mu^i\text{Nd}_{\text{source}}}}{\partial \mu^i\text{Nd}_{\text{CAI}}} \right)^2 \sigma_{\mu^i\text{Nd}_{\text{CAI}}}^2 \quad (5)$$

Where F refers to the function given in Equation (4). Equivalent equations can be written for Sm. The mass balance calculation was performed using mean Sm concentrations of 0.313 p.p.m. and 4.54 p.p.m. for Allende and the CAIs, respectively (that is, with chondritic Sm/Nd ratios for both objects). All input parameters and the resulting composition of the carbonaceous chondrite source reservoir are also given in Extended Data Table 2.

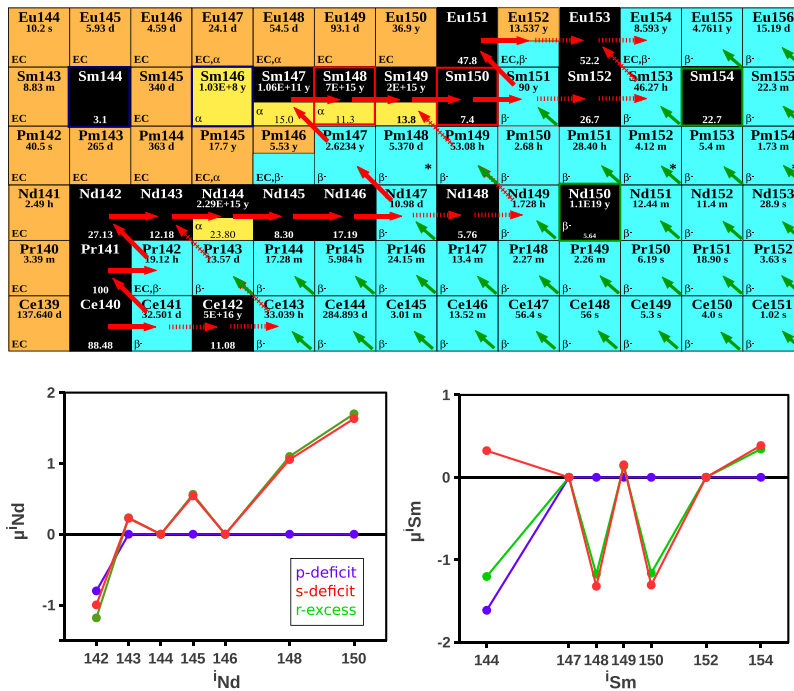
The Nd and Sm mass balance calculations indicate that the CAI-free carbonaceous chondrite source reservoir is characterized by an s-deficit relative to the Earth and the other chondrites, in both Nd and Sm isotopes. This is consistent with information derived from other isotope systems (for example, Sr, Zr, Mo, Ru) where carbonaceous chondrites are characterized by the largest s-deficits relative to the Earth, followed by ordinary and enstatite chondrites^{16,28,29,39}. We note that carbonaceous chondrite data obtained in previous studies^{15,17} also plot along the mass balance mixing relation between CAIs and a CAI-free carbonaceous chondrite source. This implies that the isotopic compositions of the other carbonaceous chondrites

are also influenced by CAI-like material, and that they derive from a common s-depleted reservoir. The fact that some of the other carbonaceous chondrites also plot on the mixing line close to the bulk Allende values, despite containing fewer CAIs than CV chondrites, might be due to the higher REE enrichments in these non-CV CAIs (for example, hibonites in CM chondrites) or the fact that CAI-like material is not present in the form of well-defined inclusions but could be dispersed in the matrix in the form of small dust grains that are partially altered by parent-body metamorphism. Because no Sm and Nd isotope data for non-CV carbonaceous chondrite CAIs are available, one can only speculate on whether or not these CAIs also might carry larger nucleosynthetic Sm and Nd anomalies than Allende CAIs.

In principle, the Nd and Sm isotope compositions observed in ordinary and enstatite chondrites could also be influenced by CAIs. However, petrographic and chemical investigations imply that CAI-like material in these chondrite types is extremely rare^{38,40–42}, and no Sm and Nd isotope data of these objects are available. Nevertheless, the effect of CAIs on the measured bulk Nd and Sm isotope composition of enstatite and ordinary chondrites is estimated to be no larger than 2 p.p.m. for Nd and 5 p.p.m. for Sm, respectively (Extended Data Table 2). This calculation assumes that the CAI-like material in ordinary and enstatite chondrites has a maximum REE enrichment of 50 times the concentrations in the CI (Ivuna-like) group of chondrites and an isotopic composition similar to normal Allende CAIs, and that the maximum CAI abundance in these chondrites is 0.05 wt.%. Given the small effects, we have omitted any correction of our measured data. However, we note that any such correction would result in slightly larger anomalies in non-radiogenic Nd isotopes and thus a higher $\mu^{142}\text{Nd}_{\text{corrected}}$; that is, an even better agreement between the nucleosynthetic anomaly-corrected $\mu^{142}\text{Nd}$ values of meteorites and the accessible Earth.

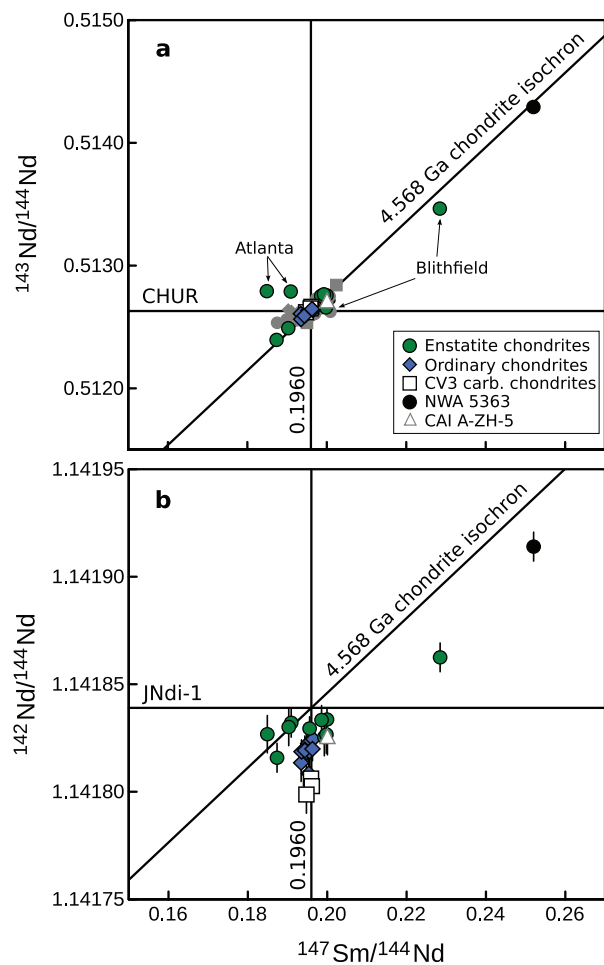
CAIs exhibit isotope anomalies in Nd and Sm, but also in many other elements^{16,18,22,30,37}. To explore the collateral effects of the mass balance between CAIs and Allende defined above for Nd and Sm on other isotope systems, we also applied it to Ca, Ti, Cr, Ni, Sr, Zr, Mo and Ba. The input parameters and results are given in Extended Data Table 4. Compared with the results from Nd and Sm, the isotopic compositions calculated for the CAI-free carbonaceous chondrite source reservoir for Ca, Ti, Cr, Ni, Sr, Zr, Mo and Ba do not differ greatly from the bulk Allende values (the most noticeable change is the reduction of the $\mu^{50}\text{Ti}$ anomaly from 365 ± 34 for bulk Allende to 221 ± 46 for the CAI-free component, consistent with the measured value (189 ± 6) of a CAI-free Allende sample¹⁸). This is explained by the fact that the chemical enrichment of these elements in the CAIs relative to the host rock are not as strong as for Nd and Sm, and that the anomalies in the CAIs and bulk Allende are less disparate than for Nd and Sm. In other words, the CAIs have less influence on the bulk Allende isotopic composition for Ca, Ti, Cr, Ni, Sr, Zr, Mo and Ba than they have for Nd and Sm. We note, however, that the calculated CAI-free Allende compositions for the Sr, Zr, and Mo isotope anomalies are fully consistent with the inferences made above from Nd and Sm—that is, the formation of the carbonaceous chondrites from a nebular reservoir that is depleted in s-process material relative to Earth.

32. Huss, G. R. Implications of isotopic anomalies and presolar grains for the formation of the Solar System. *Antarct. Meteor. Res.* **17**, 132–152 (2004).
33. Gardner-Vandy, K. G., Lauretta, D. S. & McCoy, T. J. A petrologic, thermodynamic and experimental study of brachinites: partial melt residues of an R chondrite-like precursor. *Geochim. Cosmochim. Acta* **122**, 36–57 (2013).
34. Burkhardt, C. et al. NWA 5363/NWA 5400 and the Earth: isotopic twins or just distant cousins? In *46th Lunar and Planetary Science Conference* abstr. 2732 (Lunar and Planetary Institute, 2015).
35. Burkhardt, C. et al. Hf-W mineral isochron for Ca,Al-rich inclusions: age of the solar system and the timing of core formation in planetesimals. *Geochim. Cosmochim. Acta* **72**, 6177–6197 (2008).
36. Pourmand, A., Dauphas, N. & Ireland, T. J. A novel extraction chromatography and MC-ICP-MS technique for rapid analysis of REE, Sc and Y: revising CI-chondrite and Post-Archean Australian Shale (PAAS) abundances. *Chem. Geol.* **291**, 38–54 (2012).
37. Birck, J. L. In *Geochemistry of Non-Traditional Stable Isotopes* (eds Johnson, C. M. et al.) 25–64 (Mineralogical Society of America, 2004).
38. Hezel, D. C., Russell, S. S., Ross, A. J. & Kearsley, A. T. Modal abundances of CAIs: implications for bulk chondrite element abundances and fractionations. *Meteorit. Planet. Sci.* **43**, 1879–1894 (2008).
39. Moynier, F. et al. Planetary-scale strontium isotopic heterogeneity and the age of volatile depletion of early solar system materials. *Astrophys. J.* **758**, 45 (2012).
40. Bischoff, A. & Keil, K. Al-rich objects in ordinary chondrites: related origin of carbonaceous and ordinary chondrites and their constituents. *Geochim. Cosmochim. Acta* **48**, 693–709 (1984).
41. Bischoff, A., Keil, K. & Stöffler, D. Perovskite-hibonite-spinel-bearing inclusions and Al-rich chondrules and fragments in enstatite chondrites. *Chem. Erde-Geochim.* **44**, 97–106 (1985).
42. Dauphas, N. & Pourmand, A. Thulium anomalies and rare earth element patterns in meteorites and the Earth: nebular fractionation and the nugget effect. *Geochim. Cosmochim. Acta* **163**, 234–261 (2015).
43. Nyquist, L. E. et al. ^{146}Sm – ^{142}Nd formation interval for the lunar mantle. *Geochim. Cosmochim. Acta* **59**, 2817–2837 (1995).

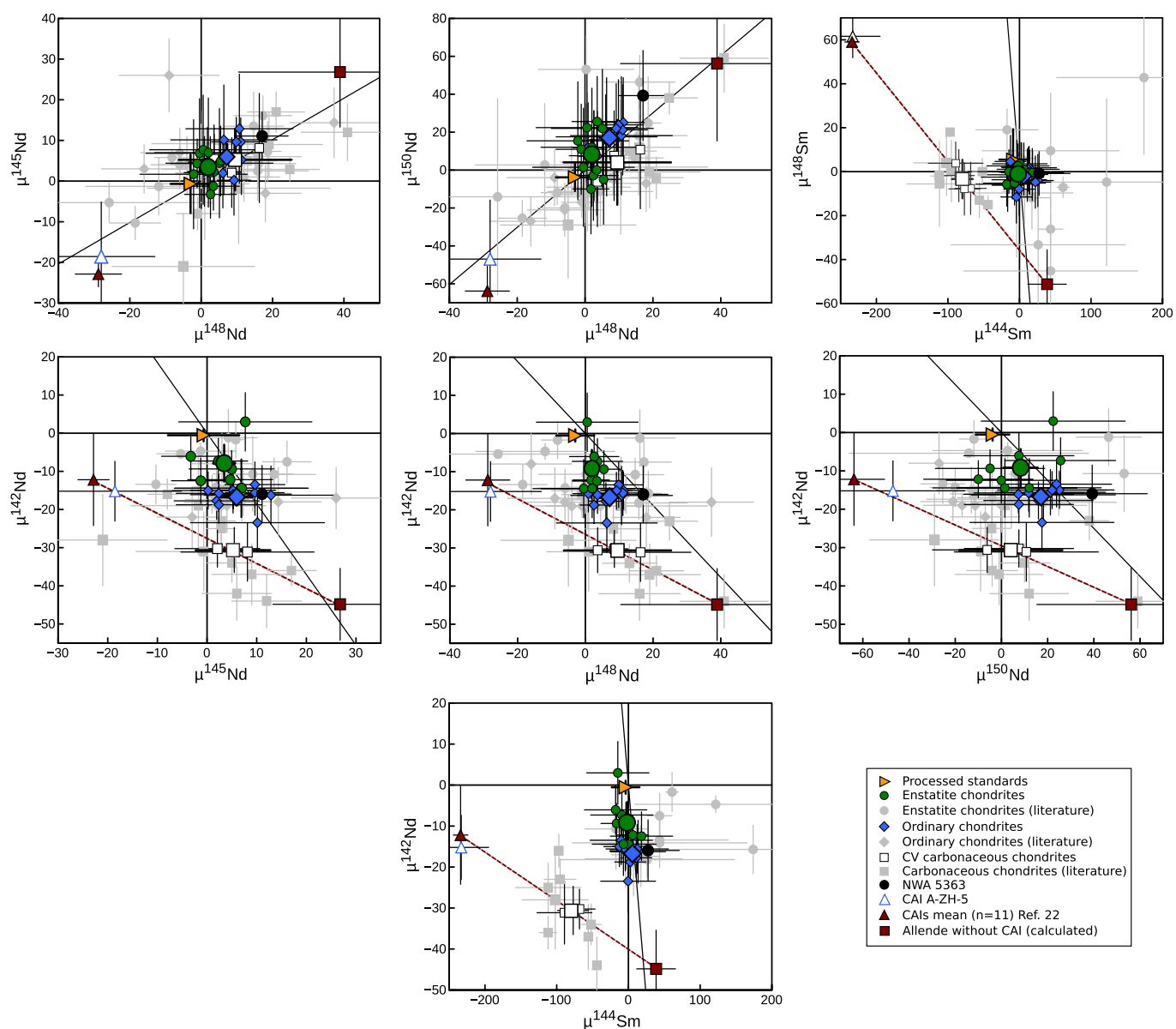


Extended Data Figure 1 | Nucleosynthetic pathways and calculated anomaly patterns for Nd and Sm. The top panel is a chart of the nuclides in the Ce–Nd–Sm–Gd mass region. Stable isotopes and their solar abundances are in black boxes on the chart, short-lived isotopes and their half-lives are in coloured boxes: blue indicates β^- unstable, orange electron capture and yellow α -decay. Solid red arrows mark the main path of s-process nucleosynthesis, the dashed red arrows mark minor

s-process branches and green arrows indicate the decay path of r-process nucleosynthesis. ^{148}Sm and ^{150}Sm are produced only by the s-process, ^{150}Nd and ^{154}Sm only by the r-process and ^{144}Sm and ^{146}Sm are p-process-only isotopes. The lower panels show expected μNd (left) and μSm (right) anomaly patterns for a p-process deficit (purple), an s-process deficit (red) and an r-process excess (green) for internal normalization to $^{146}\text{Nd}/^{144}\text{Nd}$ and $^{152}\text{Sm}/^{147}\text{Sm}$, calculated using stellar model abundances²⁷.



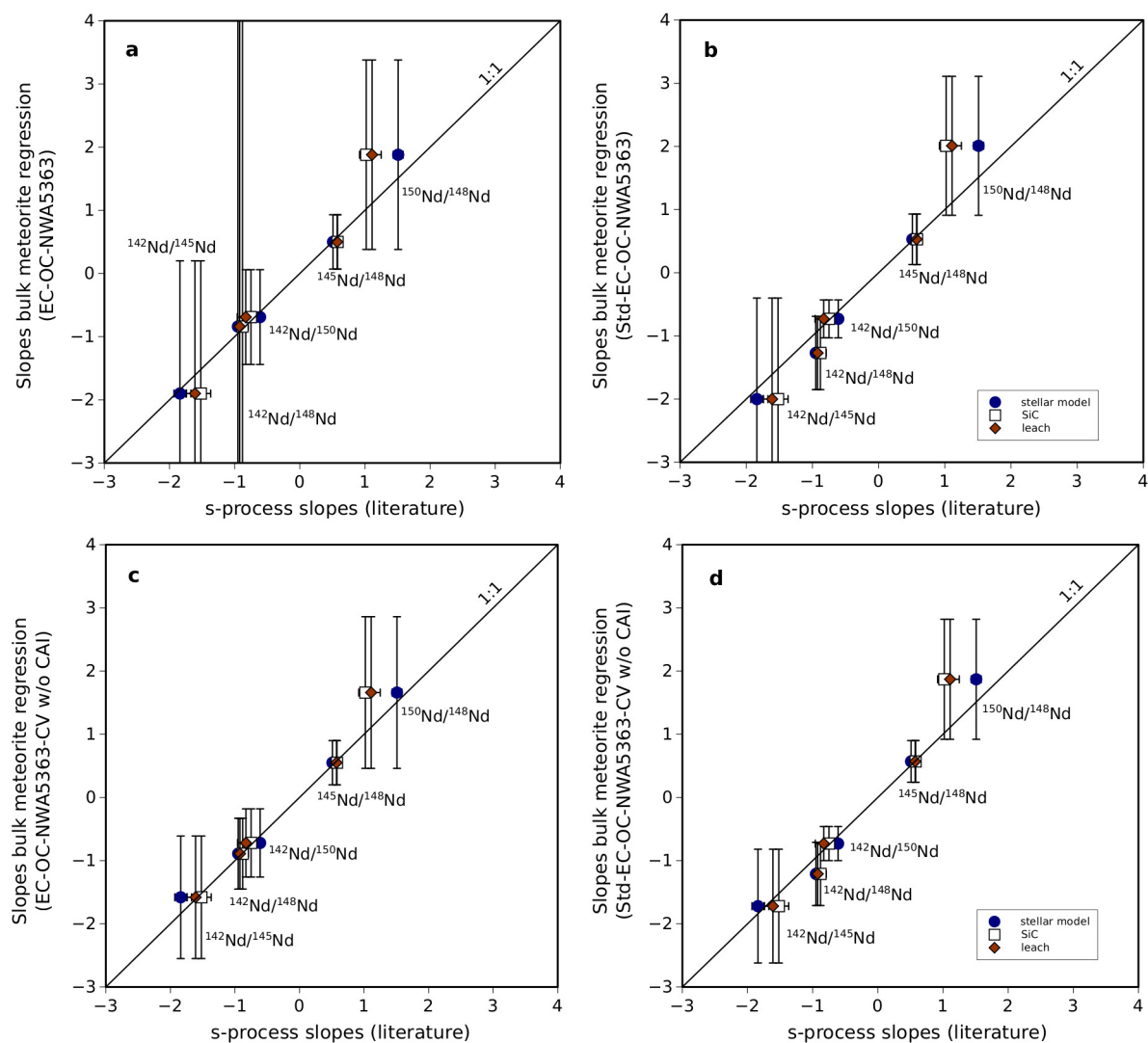
Extended Data Figure 2 | Sm/Nd isochron diagrams of measured meteorite samples. **a**, For $^{143}\text{Nd}/^{144}\text{Nd}$, all but the disturbed Atlanta and Blithfield chondrites cluster in a narrow range around a 4.568 Ga chondrite isochron, consistent with literature data (grey). **b**, For $^{142}\text{Nd}/^{144}\text{Nd}$, the meteorite data mostly fall below a 4.568 Ga isochron constructed through the accessible Earth value and only poorly correlate with Sm/Nd, indicating that, aside from Sm/Nd fractionation and ^{146}Sm decay, other processes are responsible for setting the $^{142}\text{Nd}/^{144}\text{Nd}$ of meteorites. Error bars represent the external reproducibility (2 s.d. of the standards run in the same measurement campaign as the samples).



Extended Data Figure 3 | Comparison of Nd and Sm isotope data.

The new data agree with literature data (in grey), but show less scatter, facilitating the calculation of more precise group averages. The error bars shown for our measurements represent external reproducibility (2 s.d. of the standards run in the same measurement campaign as the samples),

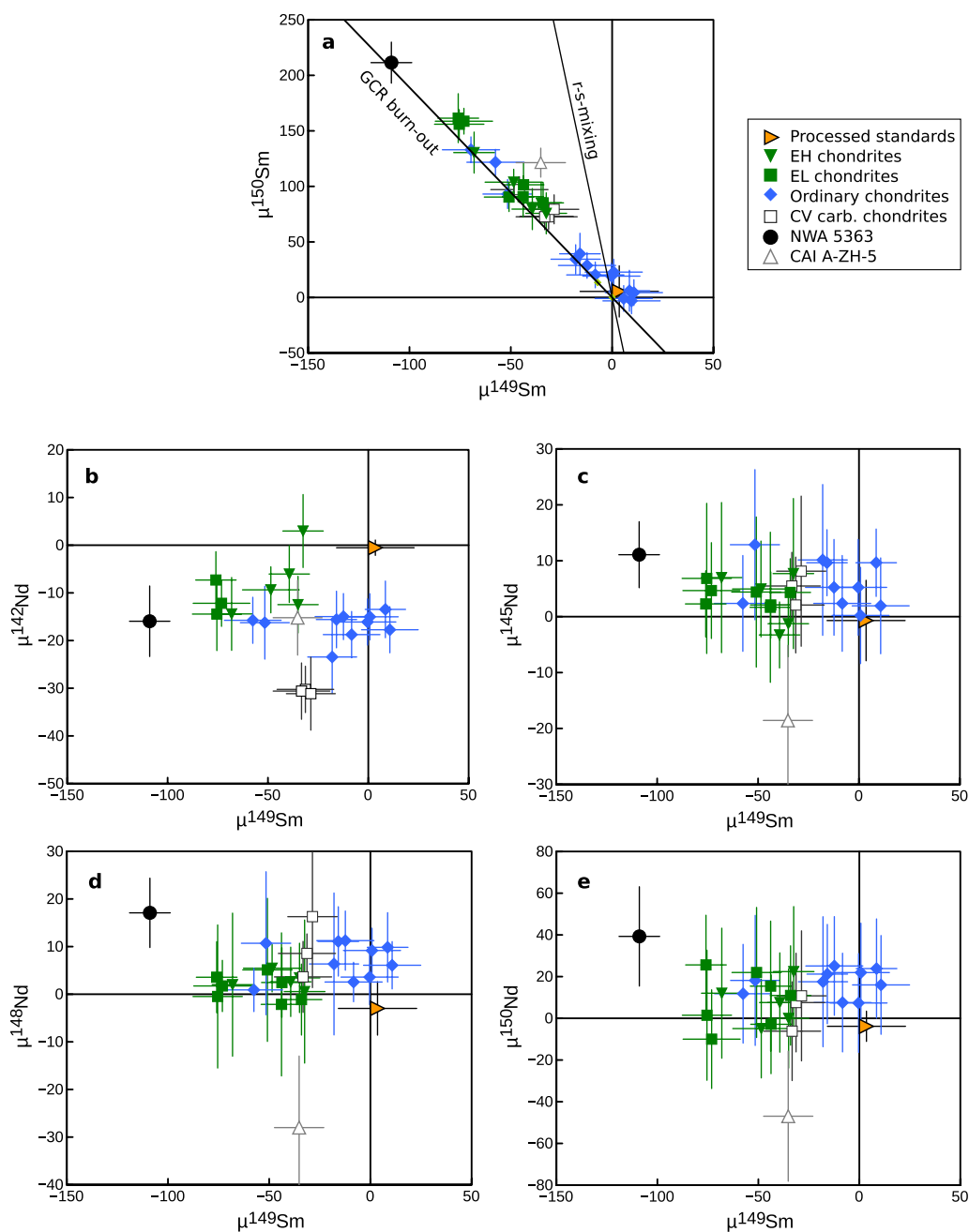
whereas the uncertainties for the literature values are the 2 s.e. of the measurements. The solid lines denote mixing of the s-model prediction²⁷ with the terrestrial composition. The dashed lines are the mixing line between CAIs and the CAI-free carbonaceous chondrite source reservoir as calculated by isotopic mass balance.



Extended Data Figure 4 | Comparison of the slopes obtained from bulk meteorite anomaly data regressions and the slopes obtained from s-process modelling, SiC grain data and chondrite leachate data.

a, Slopes from the regression of enstatite chondrite, ordinary chondrite and NWA 5363 data. **b**, The same as **a** but including the processed standard data in the regression. **c**, Slopes from the regression of enstatite chondrite, ordinary chondrite and NWA 5363 values and calculated CAI-free Allende point ('CV w/o CAI'). **d**, The same as **c** but including the processed standard data in the regression. Within uncertainties, the

slopes from the bulk meteorite regressions are indistinguishable from the slopes from the literature data^{20,21,26,27}, no matter which samples are used in the regressions. This implies that the Nd isotope variations in enstatite chondrites, ordinary chondrites, NWA 5363 and the CAI-free carbonaceous chondrite source are due to s-process heterogeneities. All regressions were performed using ISOPLOT. The slopes and $\mu^{142}\text{Nd}$ intercepts of the regressions are also given in Extended Data Table 3. Error bars are the 95% CI.



Extended Data Figure 5 | Effects of meteoroid exposure to galactic cosmic rays (GCRs) on the Sm and Nd isotope compositions.

a, Meteorites of this study show correlated $\mu^{149}\text{Sm}$ and $\mu^{150}\text{Sm}$ anomalies that are consistent with GCR exposure. Such reactions can also alter the Nd isotope signatures of planetary materials⁴³. However, given the much smaller neutron capture cross-sections of the Nd isotopes relative to ^{149}Sm ,

any effect of GCRs on $\mu^{142}\text{Nd}$ is <1 p.p.m. **b–e**, Within a given meteorite group no obvious correlations are seen in $\mu^i\text{Nd}$ versus $\mu^{149}\text{Sm}$, indicating the absence of significant GCR effects on the Nd isotope data. Error bars represent the external reproducibility (2 s.d. of the standards run in the same measurement campaign as the samples).

Extended Data Table 1 | Measured and calculated $^{147}\text{Sm}/^{144}\text{Nd}$ and $\mu^{142}\text{Nd}$ values

Sample	Type	$^{147}\text{Sm}/^{144}\text{Nd}$ measured	2se	$\mu^{142}\text{Nd}$ measured	2sd	$\mu^{142}\text{Nd}$ corrected 1	2sd	$^{143}\text{Nd}/^{144}\text{Nd}$ measured	2sd	$^{147}\text{Sm}/^{144}\text{Nd}$ calculated	2sd	$\mu^{142}\text{Nd}$ corrected 2	2sd
Hvittis (1)	EL6	0.1999	0.0002	-6	5	-12	5	0.5127579	0.0000027	0.2003	0.0001	-13	5
Hvittis (2)	EL6	0.1986	0.0002	-3	6	-7	6	0.5127533	0.0000060	0.2001	0.0002	-10	6
Hvittis (3)	EL6	0.1993	0.0002	-10	8	-14	8	0.5127663	0.0000051	0.2005	0.0002	-16	8
Atlanta (1)	EL6	<i>0.1909</i>	<i>0.0002</i>	<i>-5</i>	<i>6</i>	<i>3</i>	<i>6</i>	<i>0.5127888</i>	<i>0.0000060</i>	<i>0.2013</i>	<i>0.0002</i>	<i>-12</i>	<i>6</i>
Atlanta (2)	EL6	<i>0.1849</i>	<i>0.0002</i>	<i>-8</i>	<i>8</i>	<i>8</i>	<i>8</i>	<i>0.5127919</i>	<i>0.0000051</i>	<i>0.2014</i>	<i>0.0002</i>	<i>-16</i>	<i>8</i>
Blithfield (1)	EL6	<i>0.2285</i>	<i>0.0002</i>	<i>22</i>	<i>6</i>	<i>-26</i>	<i>6</i>	<i>0.5134645</i>	<i>0.0000060</i>	<i>0.2236</i>	<i>0.0002</i>	<i>-19</i>	<i>6</i>
Blithfield (2)	EL6	<i>0.1998</i>	<i>0.0002</i>	<i>-9</i>	<i>8</i>	<i>-14</i>	<i>8</i>	<i>0.5126591</i>	<i>0.0000051</i>	<i>0.1970</i>	<i>0.0002</i>	<i>-10</i>	<i>8</i>
St. Sauveur	EH6	0.1956	0.0002	-10	5	-9	5	0.5126239	0.0000027	0.1958	0.0001	-10	5
Abee (1)	EH4	0.1874	0.0002	-19	6	-6	6	0.5123947	0.0000060	0.1883	0.0002	-7	6
Abee (2)	EH4	0.1903	0.0002	-5	8	3	8	0.5124901	0.0000051	0.1914	0.0002	1	8
Indarch (1)	EH4	0.1953	0.0002	-14	6	-12	6	0.5126219	0.0000060	0.1958	0.0002	-13	6
Indarch (2)	EH4	0.1948	0.0002	-16	8	-14	8	0.5126109	0.0000051	0.1954	0.0002	-15	8
Av. enstatite chondrites				-10.4	4.5	-9.2	4.9					-10.4	3.4
Queens Mercy	H6	0.1946	0.0002	-20	5	-18	5	0.5125971	0.0000027	0.1950	0.0001	-18	5
Allegan	H5	0.1952	0.0002	-16	5	-15	5	0.5126148	0.0000027	0.1955	0.0001	-15	5
Forest City	H5	0.1944	0.0002	-19	5	-16	5	0.5125989	0.0000027	0.1950	0.0001	-17	5
Rultusk	H5	0.1934	0.0002	-20	8	-16	8	0.5126079	0.0000051	0.1953	0.0002	-19	8
Ste. Marguerite (1)	H4	0.1955	0.0002	-16	6	-16	6	0.5126351	0.0000060	0.1962	0.0002	-17	6
Ste. Marguerite (2)	H4	0.1954	0.0002	-24	8	-23	8	0.5126355	0.0000051	0.1962	0.0002	-25	8
Bruderheim	L6	0.1935	0.0002	-19	5	-16	5	0.5125629	0.0000027	0.1938	0.0001	-16	5
Farmington (2)	L5	0.1944	0.0002	-16	6	-13	6	0.5125907	0.0000060	0.1947	0.0002	-14	6
Dhumsala	LL6	0.1965	0.0002	-14	5	-15	5	0.5126368	0.0000027	0.1963	0.0001	-15	5
Chelyabinsk	LL5	0.1963	0.0002	-18	5	-19	5	0.5126469	0.0000027	0.1966	0.0001	-19	5
Av. ordinary chondrites				-18.3	2.1	-16.7	2.0					-17.5	2.2
Allende (2)	CV3	0.1959	0.0002	-30	5	-30	5	0.5126511	0.0000027	0.1967	0.0001	-31	5
Allende (3)	CV3	0.1961	0.0002	-30	6	-31	6	0.5126644	0.0000060	0.1972	0.0002	-32	6
Allende (4)	CV3	0.1948	0.0002	-33	8	-31	8	0.5126204	0.0000051	0.1957	0.0002	-33	8
Average CV				-31.3	3.7	-30.7	1.1					-32.1	1.4
NWA 5363	Ung.	0.2520	0.0005	67.1	5.9	-16.0	7.5	0.5142920	0.0000060	0.2509	0.0002	-14.2	7.4
AZH-5	CAI	0.2000	0.0012	-9.2	7.6	-15.2	7.8	0.5127164	0.0000051	0.1989	0.0002	-13.5	7.7

To investigate the effect of nucleosynthetic anomalies on $\mu^{142}\text{Nd}$ with high precision, the measured $\mu^{142}\text{Nd}$ values of the meteorites first need to be corrected for ^{146}Sm decay to a constant $^{147}\text{Sm}/^{144}\text{Nd} = 0.1960$ (ref. 1) and assuming a common 4.568 Gy evolution with an initial Solar System value of $^{146}\text{Sm}/^{144}\text{Sm} = 0.00828 \pm 0.00044$ (ref. 23). This can be done either by using the measured $^{147}\text{Sm}/^{144}\text{Nd}$ values ($\mu^{142}\text{Nd}$ corrected 1'), or the $^{147}\text{Sm}/^{144}\text{Nd}$ values are first calculated from the measured $^{143}\text{Nd}/^{144}\text{Nd}$, a chondritic $^{143}\text{Nd}/^{144}\text{Nd} = 0.512630$ (ref. 1) and the decay constant $\lambda^{147}\text{Sm} = 6.539 \times 10^{-12} \text{ yr}^{-1}$ ($\mu^{142}\text{Nd}$ corrected 2'). The latter method is insensitive to recent changes in the Sm/Nd ratio (through terrestrial weathering or incomplete spike-sample equilibrium, for example) whereas the former is less model-dependent. Within uncertainties, both correction methods yield indistinguishable $\mu^{142}\text{Nd}$ values and with the exception of Abee and the radiogenic NWA 5363, these values are also indistinguishable from the measured values. For both corrections the uncertainties on the initial $^{146}\text{Sm}/^{144}\text{Sm}$, $^{147}\text{Sm}/^{144}\text{Nd}$ and the measured $\mu^{142}\text{Nd}$ were propagated, but a significant change is only observed for NWA 5363, whose decay correction (83 p.p.m.) changed the uncertainty from ± 6 p.p.m. to ± 7.5 p.p.m. Data for the Atlanta and Blithfield EL6 chondrites are excluded (italic) owing to their disturbed Sm/Nd systematics.

Extended Data Table 2 | Input parameters and the results of isotopic mass balance calculations for Nd and Sm

	Mass balance Allende – CAIs																
	Nd (ppm)	Sm (ppm)	$^{147}\text{Sm}/^{144}\text{Nd}$	$\mu^{142}\text{Nd}$	2σ	$\mu^{145}\text{Nd}$	2σ	$\mu^{148}\text{Nd}$	2σ	$\mu^{150}\text{Nd}$	2σ	$\mu^{144}\text{Sm}$	2σ	$\mu^{148}\text{Sm}$	2σ	$\mu^{154}\text{Sm}$	2σ
CAI	14	4.54	0.1960	-12	12	-23	3	-29	7	-64	13	-234	10	59	3	-18	6
Allende	0.967	0.313	0.1960	-31	1	5	8	9	16	4	22	-80	15	-3	9	-7	13
CAI fraction = 0.03																	
Allende w/o CAI	0.564	0.183	0.1960	-45	9	27	14	39	28	56	41	39	27	-51	16	1	23

	Mass balance enstatite chondrites – CAIs																
	Nd (ppm)	Sm (ppm)	$^{147}\text{Sm}/^{144}\text{Nd}$	$\mu^{142}\text{Nd}$	2σ	$\mu^{145}\text{Nd}$	2σ	$\mu^{148}\text{Nd}$	2σ	$\mu^{150}\text{Nd}$	2σ	$\mu^{144}\text{Sm}$	2σ	$\mu^{148}\text{Sm}$	2σ	$\mu^{154}\text{Sm}$	2σ
CAI	25	8.10	0.1960	-12	12	-23	3	-29	7	-64	13	-234	10	59	3	-18	6
ECs	0.486	0.157	0.1960	-9	5	3	2	2	2	8	7	-2	8	-1	2	0	4
CAI fraction = 0.005																	
ECs w/o CAI	0.474	0.153	0.1960	-9	5	4	2	3	2	10	8	4	8	-3	2	1	4

	Mass balance ordinary chondrites – CAIs																
	Nd (ppm)	Sm (ppm)	$^{147}\text{Sm}/^{144}\text{Nd}$	$\mu^{142}\text{Nd}$	2σ	$\mu^{145}\text{Nd}$	2σ	$\mu^{148}\text{Nd}$	2σ	$\mu^{150}\text{Nd}$	2σ	$\mu^{144}\text{Sm}$	2σ	$\mu^{148}\text{Sm}$	2σ	$\mu^{154}\text{Sm}$	2σ
CAI	25	8.10	0.1960	-12	12	-23	3	-29	7	-64	13	-234	10	59	3	-18	6
OCs	0.680	0.220	0.1960	-17	2	6	3	7	3	17	5	6	7	-2	3	1	4
CAI fraction = 0.005																	
OCs w/o CAI	0.668	0.216	0.1960	-17	2	7	3	8	3	19	5	10	8	-3	3	1	5

Uncertainties for CAIs, Allende, as well as enstatite and ordinary chondrites, represent the two-sided Student's *t*-test 95% CI and were propagated throughout the mass balance calculation according to equation (5).

Extended Data Table 3 | $\mu^{142}\text{Nd}$ values corrected for nucleosynthetic anomalies

(a) Nucleosynthetic anomaly corrected $\mu^{142}\text{Nd}$ in parts per million deviation relative to the mean measured JNd-1 standard													
regression EC, OC, NWA							regression Std, EC, OC, NWA						
correction from intercept of	^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	wt. av. 2σ
correction relation	^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	wt. av. 2σ
slope rel $\mu^{142}\text{Nd}$	-1.9	2.1	-0.84	9.8	-0.69	0.75	-2	1.6	-1.27	0.58	-0.73	0.3	
$\mu^{142}\text{Nd}_{\text{anomaly corrected}}$	-3	12	-9	74	-4	12	-3	9	-6	4	-3	4	-4 3
(a) continued Nucleosynthetic anomaly corrected $\mu^{142}\text{Nd}$ in parts per million deviation relative to the mean measured JNd-1 standard													
regression EC, OC, NWA, CV w/o CAI							regression Std, EC, OC, NWA, CV w/o CAI						
correction from intercept of	^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	wt. av. 2σ
correction relation	^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	wt. av. 2σ
slope rel $\mu^{142}\text{Nd}$	-1.58	0.97	-0.89	0.56	-0.72	0.54	-1.72	0.9	-1.21	0.5	-0.73	0.27	
$\mu^{142}\text{Nd}_{\text{anomaly corrected}}$	-5	7	-9	5	-4	9	-4	6	-6	4	-3	4	-5 3
(b) Nucleosynthetic anomaly corrected $\mu^{142}\text{Nd}$ in parts per million deviation relative to the mean measured JNd-1 standard													
stellar model							SiC						
correction using	^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	wt. av. 2σ
correction relation	^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	wt. av. 2σ
slope rel $\mu^{142}\text{Nd}$	-1.84		-0.95		-0.61		-1.52		-0.88		-0.75		
EC $\mu^{142}\text{Nd}_{\text{anomaly corrected}}$	-3	6	-7	5	-4	7	-4	6	-8	5	-3	7	-5 3
OC $\mu^{142}\text{Nd}_{\text{anomaly corrected}}$	-6	6	-10	3	-6	3	-8	5	-10	3	-4	4	-8 2
NWA $\mu^{142}\text{Nd}_{\text{anomaly corrected}}$	4	13	0	10	8	16	1	12	-1	10	14	19	3 7
wt. av. $\mu^{142}\text{Nd}_{\text{anomaly corrected}}$													-5 2
(b) continued Nucleosynthetic anomaly corrected $\mu^{142}\text{Nd}$ in parts per million deviation relative to the mean measured JNd-1 standard													
leachates							leachates						
^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	wt. av. 2σ	^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	wt. av. 2σ
-1.61		-0.92		-0.83			-1.61		-0.92		-0.83		
-4	5	-7	9	-2	5	-4 3	-4	5	-7	9	-2	5	-4 3
-7	5	-10	3	-3	4	-7 2	-7	5	-10	3	-3	4	-7 2
2	12	0	10	17	21	4 8	2	12	0	10	17	21	4 8
													-4 2
(a') Nucleosynthetic anomaly corrected $\mu^{142}\text{Nd}$ in parts per million deviation relative to the mean of processed standards													
regression EC, OC, NWA							regression Std, EC, OC, NWA						
correction from intercept of	^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	wt. av. 2σ
correction relation	^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	wt. av. 2σ
slope rel $\mu^{142}\text{Nd}$	-1.9	2.1	-0.84	9.8	-0.69	0.75	-2	1.6	-1.27	0.58	-0.73	0.3	
$\mu^{142}\text{Nd}_{\text{anomaly corrected}}$	-1	14	-6	100	-1	15	-1	10	-2	5	0	5	-1 4
(a') continued Nucleosynthetic anomaly corrected $\mu^{142}\text{Nd}$ in parts per million deviation relative to the mean of processed standards													
regression EC, OC, NWA, CV w/o CAI							regression Std, EC, OC, NWA, CV w/o CAI						
correction from intercept of	^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	wt. av. 2σ
correction relation	^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	wt. av. 2σ
slope rel $\mu^{142}\text{Nd}$	-1.58	0.97	-0.89	0.56	-0.72	0.54	-1.72	0.9	-1.21	0.5	-0.73	0.27	
$\mu^{142}\text{Nd}_{\text{anomaly corrected}}$	-3	8	-6	6	0	11	-2	7	-2	5	0	5	-1 3
(b') Nucleosynthetic anomaly corrected $\mu^{142}\text{Nd}$ in parts per million deviation relative to the mean of processed terrestrial standards													
stellar model							SiC						
correction using	^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	wt. av. 2σ
correction relation	^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	^{145}Nd	2σ	^{148}Nd	2σ	^{150}Nd	2σ	wt. av. 2σ
slope rel $\mu^{142}\text{Nd}$	-1.84		-0.95		-0.61		-1.52		-0.88		-0.75		
EC $\mu^{142}\text{Nd}_{\text{anomaly corrected}}$	-1	6	-4	5	-1	7	-2	6	-4	5	1	7	-2 3
OC $\mu^{142}\text{Nd}_{\text{anomaly corrected}}$	-4	6	-7	3	-3	3	-6	5	-7	3	-1	4	-5 2
NWA $\mu^{142}\text{Nd}_{\text{anomaly corrected}}$	6	13	4	10	11	16	3	12	2	10	17	19	6 7
wt. av. $\mu^{142}\text{Nd}_{\text{anomaly corrected}}$													-2 2

a. Correction obtained from the intercept values of regressions through the measured meteorite Nd isotope data in $\mu^{142}\text{Nd}$ versus $\mu^{142}\text{Nd}$ (where $i = 145, 148$ and 150) space (see also Fig. 3; Extended Data Fig. 4). **b.** Correction calculated from the intercepts of the measured data points and the slopes of the s-process modelling²⁷, isotopic compositions of SiC grains²⁶ and isotopic compositions of chondrite leachates^{20,21} using the equation $\mu^{142}\text{Nd}_{\text{anomaly corrected}} = \mu^{142}\text{Nd} - \mu^{142}\text{Nd} \times \text{slope}$. EC, enstatite chondrites; OC, ordinary chondrites; NWA, NWA 5363; Std, processed terrestrial standards; CV without CAI, CAI-free Allende component as calculated from isotopic mass balance. Regressions were calculated using ISOPLLOT and uncertainties of the intercept value are the 95% CI. All anomaly-corrected $\mu^{142}\text{Nd}$ values calculated for the individual meteorites are indistinguishable within uncertainty, regardless of the technique used to make the corrections (that is, using regressions through the bulk meteorite Nd data, s-process model predictions, SiC grain data or acid leachate data). The weighted averages of the anomaly-corrected $\mu^{142}\text{Nd}$ values consistently range between -6 ± 4 p.p.m. and -4 ± 2 p.p.m. relative to the mean measured JNd-1 standard value. If all data are normalized to the mean values measured for the processed standards (**a'**, **b'**), the anomaly-corrected $\mu^{142}\text{Nd}$ values range between -4 ± 5 p.p.m. and -1 ± 2 p.p.m.

Extended Data Table 4 | Collateral effects of the isotopic mass balance between Allende and CAIs for Ca, Ti, Cr, Ni, Sr, Zr, Mo and Ba

CAI fraction = 0.03	Ca (wt%)	$\mu^{46}\text{Ca}$	2σ
CAI	10.1	370	160
Allende	1.9	392	50
Allende w/o CAI	1.6	396	67

CAI fraction = 0.03	Ti (ppm)	$\mu^{46}\text{Ti}$	2σ	$\mu^{50}\text{Ti}$	2σ
CAI	6042	172	12	933	69
Allende	899	67	7	365	34
Allende w/o CAI	739	40	9	221	46

CAI fraction = 0.03	Cr (ppm)	$\mu^{54}\text{Cr}$	2σ
CAI	997	641	90
Allende	3638	87	7
Allende w/o CAI	3720	82	7

CAI fraction = 0.03	Ni (ppm)	$\mu^{62}\text{Ni}$	2σ	$\mu^{64}\text{Ni}$	2σ
CAI	342	117	20	247	58
Allende	14193	11	3	31	9
Allende w/o CAI	14621	11	3	31	9

CAI fraction = 0.03	Sr (ppm)	$\mu^{84}\text{Sr}$	2σ
CAI	66	126	11
Allende	16	63	10
Allende w/o CAI	14	54	12

CAI fraction = 0.03	Zr (ppm)	$\mu^{91}\text{Zr}$	2σ	$\mu^{92}\text{Zr}$	2σ	$\mu^{96}\text{Zr}$	2σ
CAI	40	0	6	-2	14	161	31
Allende	7	2	21	-3	8	110	31
Allende w/o CAI	6	2	26	-3	10	99	38

CAI fraction = 0.03	Mo (ppm)	$\mu^{92}\text{Mo}$	2σ	$\mu^{94}\text{Mo}$	2σ	$\mu^{95}\text{Mo}$	2σ	$\mu^{97}\text{Mo}$	2σ	$\mu^{100}\text{Mo}$	2σ
CAI	3.5	274	21	123	19	197	8	89	7	131	22
Allende	1.5	287	67	210	51	168	34	94	43	100	48
Allende w/o CAI	1.4	288	72	217	55	166	37	94	47	98	52

CAI fraction = 0.03	Ba (ppm)	$\mu^{130}\text{Ba}$	2σ	$\mu^{132}\text{Ba}$	2σ	$\mu^{135}\text{Ba}$	2σ	$\mu^{137}\text{Ba}$	2σ	$\mu^{138}\text{Ba}$	2σ
CAI	30	-40	44	-119	74	54	6	18	5	17	9
Allende	5	63	130	13	258	26	41	19	25	9	32
Allende w/o CAI	4	87	161	44	318	20	50	19	31	8	39

Uncertainties represent the two-sided Student's *t*-test 95% CI and were propagated throughout the mass balance calculation according to equation (5). Data from refs 15–18, 22, 28–31, 37, 39 and therein.

Primitive Solar System materials and Earth share a common initial ^{142}Nd abundance

A. Bouvier¹ & M. Boyet²

The early evolution of planetesimals and planets can be constrained using variations in the abundance of neodymium-142 (^{142}Nd), which arise from the initial distribution of ^{142}Nd within the protoplanetary disk and the radioactive decay of the short-lived samarium-146 isotope (^{146}Sm)^{1,2}. The apparent offset in ^{142}Nd abundance found previously between chondritic meteorites and Earth^{1,2} has been interpreted either as a possible consequence of nucleosynthetic variations within the protoplanetary disk^{2–4} or as a function of the differentiation of Earth very early in its history⁵. Here we report high-precision Sm and Nd stable and radiogenic isotopic compositions of four calcium–aluminium-rich refractory inclusions (CAIs) from three CV-type carbonaceous chondrites, and of three whole-rock samples of unequilibrated enstatite chondrites. The CAIs, which are the first solids formed by condensation from the nebular gas, provide the best constraints for the isotopic evolution of the early Solar System. Using the mineral isochron method for individual CAIs, we find that CAIs without isotopic anomalies in Nd compared to the terrestrial composition share a $^{146}\text{Sm}/^{144}\text{Sm}$ – $^{142}\text{Nd}/^{144}\text{Nd}$ isotopic evolution with Earth. The average $^{142}\text{Nd}/^{144}\text{Nd}$ composition for pristine enstatite chondrites that we calculate coincides with that of the accessible silicate layers of Earth. This relationship between CAIs, enstatite chondrites and Earth can only be a result of Earth having inherited the same initial abundance of ^{142}Nd and chondritic proportions of Sm and Nd. Consequently, ^{142}Nd isotopic heterogeneities found in other CAIs and among chondrite groups may arise from extrasolar grains that were present in the disk and incorporated in different proportions into these planetary objects. Our finding supports a chondritic Sm/Nd ratio for the bulk silicate Earth and, as a consequence, chondritic abundances for other refractory elements. It also removes the need for a hidden reservoir or for collisional erosion scenarios^{5,6} to explain the $^{142}\text{Nd}/^{144}\text{Nd}$ composition of Earth.

The ^{146}Sm – ^{142}Nd short-lived radiometric pair records the first few hundred million years (Myr) of the Solar System and provides a powerful geochemical tool for tracing the early silicate differentiation of planetary objects. However, its use as a precise chronometer has become more problematic in the past few years, owing to uncertainties in the half-life of ^{146}Sm (ref. 7), in the initial abundance of ^{146}Sm (ref. 8) and in the bulk Sm/Nd ratios of planetary bodies⁹. In addition, $^{142}\text{Nd}/^{144}\text{Nd}$ ratios in chondrites deviate by up to -40 p.p.m. from the terrestrial value^{2–4}, and model $^{146}\text{Sm}/^{144}\text{Sm}$ – $^{142}\text{Nd}/^{144}\text{Nd}$ ages for silicate differentiation depend on the initial $^{142}\text{Nd}/^{144}\text{Nd}$ composition of the planets from which internal reservoirs subsequently evolved. The deviations of sample compositions from the Nd isotopic standard (and similarly for Sm normalized to ^{152}Sm) for a given isotope of mass i are expressed (in parts per million) as $\mu^i\text{Nd} = [(^{i}\text{Nd}/^{144}\text{Nd})_{\text{sample}} / (^{i}\text{Nd}/^{144}\text{Nd})_{\text{reference}} - 1] \times 10^6$.

Isotopic measurements of planetary materials suggest that the solar nebula was not completely homogenized during the period of planetary accretion¹⁰. Large isotope anomalies have been detected at

the microscale for presolar grains¹¹, and smaller anomalies are found for many refractory elements at the whole-rock scale (for example, Cr, Ti (ref. 12), Sr (ref. 13), Sm and Nd (ref. 3)). Among the different chondrite groups, the carbonaceous chondrites have stable isotopic compositions for several refractory elements furthest from the

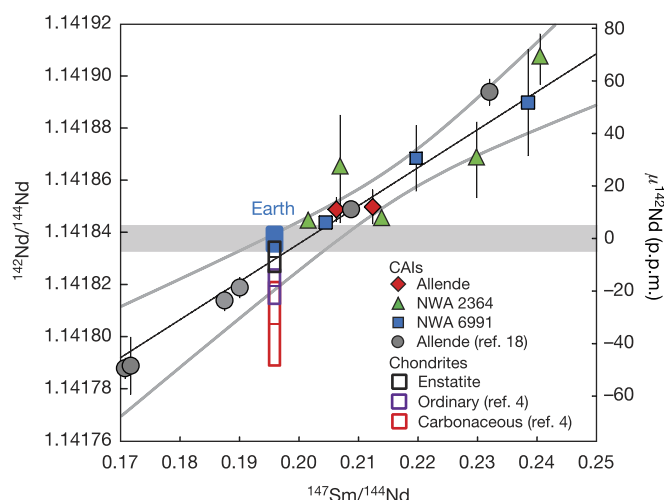


Figure 1 | ^{146}Sm – ^{142}Nd mineral and bulk isochron of the Allende, NWA 2364 and NWA 6991 CAIs compared with the Allende Al3S4 CAI, the modern terrestrial composition and bulk-rock chondrite averages for enstatite, ordinary and carbonaceous chondrites. The data for Allende Al3S4 CAI are from ref. 18 and those for ordinary and carbonaceous chondrites are from ref. 4; all other data are from this study. The modern terrestrial composition is indicated by the blue rectangle labelled Earth, and is also represented by the grey band for the average $^{142}\text{Nd}/^{144}\text{Nd}$ ratio, which corresponds to $\mu^{142}\text{Nd}$ of the JNdi-1 standard (0 ± 5 p.p.m.; 2 s.d.). Regressions of NWA 6991 (black line) and NWA 2364 (not shown) yield $^{146}\text{Sm}/^{144}\text{Sm}$ ratios of 0.0072 ± 0.0024 (MSWD = 0.76) and 0.0069 ± 0.0056 (MSWD = 2.8), respectively, at the age of Sm–Nd isotopic closure. Considering data from all CAIs from this study, the slope of the regression line (not shown) is 0.0073 ± 0.0022 (MSWD = 3.3). The NWA 6991 CAI does not have any stable isotopic variations in Nd relative to Earth, removing the need for potential corrections from nucleosynthetic or neutron capture effects. The internal isochron regression of NWA 6991 (black line; grey lines indicate the 95% confidence interval) intersects the $^{142}\text{Nd}/^{144}\text{Nd}$ compositions of Earth and enstatite chondrites at a common $^{147}\text{Sm}/^{144}\text{Nd}$ ratio, which we find matches the chondritic ratio of 0.196 (ref. 27). All errors bars are two standard errors (2 s.e.) for individual data points. Respective means and errors for chondrite groups are given in the text; those for samples and standards from this study are provided in Supplementary Table 1. The left y axis shows the $^{142}\text{Nd}/^{144}\text{Nd}$ ratios; the right y axis shows the $\mu^{142}\text{Nd}$ values (in parts per million; deviation relative to the mean obtained on the JNdi-1 terrestrial standard). The heights of the rectangles for the chondrite groups indicate 2 s.d. on the $\mu^{142}\text{Nd}$ averages (horizontal bars); the widths indicate 2 s.e. on the CHUR $^{147}\text{Sm}/^{144}\text{Nd}$ ratio²⁷.

¹University of Western Ontario, Department of Earth Sciences, Centre for Planetary Science and Exploration, London, Ontario N6A 3K7, Canada. ²Clermont Université, Université Blaise Pascal, Laboratoire Magmas et Volcans, UMR CNRS 6524, Campus Universitaire des Cézeaux, 6 avenue Blaise Pascal, 63178 Aubière Cedex, France.

terrestrial composition, and so must constitute a small mass fraction of Earth's building blocks¹⁴. The enstatite chondrites and a few ungrouped chondrites and achondrite meteorites are the only meteorite groups so far that share identical oxygen isotopic compositions¹⁵ with Earth and the Moon, suggesting a common reservoir for their accretion. The ordinary chondrites have an average $^{142}\text{Nd}/^{144}\text{Nd}$ deficit of -19 ± 5 p.p.m. (where the error indicates 2 standard deviations (s.d.); ref. 4) relative to modern terrestrial samples, whereas $^{142}\text{Nd}/^{144}\text{Nd}$ ratios in enstatite chondrites vary between those found in ordinary chondrites and those found in Earth^{2–4}.

We report high-precision Sm and Nd isotope measurements and Sm/Nd ratios of some of the earliest-formed objects in the Solar System. CAIs condensed from the solar nebula gas are the oldest dated materials in the Solar System¹⁶. The internal mineral isochron method has been applied to individual CAIs to determine the initial Solar System abundances of short-lived radionuclides for several radiogenic systems¹⁷, including ^{146}Sm – ^{142}Nd (ref. 18). In the Allende CV3 carbonaceous chondrite, CAIs can be affected by thermal metamorphism^{18,19}. To avoid isotopic disturbances from metamorphism and to evaluate the extent of isotopic heterogeneities in CAIs, we selected inclusions that have previously been chemically and/or petrographically characterized and that have different crystallization histories (Methods). We investigated mineral separates (melilite and fassaite—a Ca, Al, Mg-rich silicate and an Al-rich pyroxene, respectively) and bulk sample powders of 1-cm-sized CAIs from the CV3-chondrite meteorites Northwest Africa (NWA) 2364 (the ‘crucible’, type B) and NWA 6991 (B4, compact type A; Extended Data Fig. 1) to constrain their internal Sm–Nd isotopic evolution. We also analysed as bulk powders two Allende fine-grained CAIs (Extended Data Fig. 1) that, on the basis of their fine-grained mineral textures, have not been melted since they condensed from the nebular gas. Primitive enstatite chondrites from petrologic type 3 of the EH (ALHA77295) and EL (MAC 02837 and MAC 02839) subgroups were also selected. The sulfide assemblages of these three enstatite chondrites indicate that they are the most pristine and unmetamorphosed enstatite chondrites available from the two subgroups^{4,20}, suggesting that they are the best candidates to preserve both their Sm and Nd elemental abundances and any isotopic heterogeneities inherited from their formation region in the solar nebula. For each sample, the abundances of Sm and Nd stable isotopes, including the minor proton-rich isotope ^{144}Sm (3.1% of total Sm), were measured with a precision of a few parts per million, along with the corresponding Sm/Nd elemental ratios (Supplementary Tables 1–3). Any modifications due to exposure to galactic cosmic rays were monitored, but this effect is small and does not induce any substantial modification of Nd isotope ratios (maximum 1 p.p.m.; Methods).

The two CAIs from NWA 2364 and NWA 6991 have ^{147}Sm – ^{143}Nd ages in agreement with those obtained by U–Pb radiometric dating (ref. 16), showing that the Sm–Nd chronometers have not been disturbed by secondary processes (Extended Data Fig. 7a, b). Regression of $^{147}\text{Sm}/^{144}\text{Nd}$ – $^{142}\text{Nd}/^{144}\text{Nd}$ (Fig. 1) for NWA 6991 fractions yields a $\mu^{142}\text{Nd}$ value of -8.4 ± 5.1 p.p.m. at a chondritic Sm/Nd ratio (mean square weighted deviation, MSWD = 0.76)—a value that is indistinguishable within errors from the terrestrial equivalent (0 ± 5 p.p.m.; 2 s.d., measured on JNdi-1 Nd isotopic standard). The NWA 2364 CAI regression is less well constrained (MSWD = 2.8), yielding a $\mu^{142}\text{Nd}$ value of -2.9 ± 25.4 p.p.m. (2 s.d.) at a chondritic Sm/Nd ratio. Therefore, internal isochron regression lines of two individual CAIs intersect the composition of the modern accessible Earth at a chondritic Sm/Nd ratio, which is the same as the Sm/Nd ratio of enstatite chondrites measured in this study (Fig. 1). These similar $\mu^{142}\text{Nd}$ values at chondritic Sm/Nd ratios for two CAIs are in contrast with the literature average values for enstatite, ordinary and carbonaceous chondrites, which are, in that order, increasingly deficient in ^{142}Nd (refs 2–4; Fig. 1). The CAIs have Sm isotope patterns with deficits in ^{144}Sm , ^{149}Sm and ^{154}Sm , and excesses in ^{148}Sm and ^{150}Sm (Fig. 2a). The large deficits in ^{144}Sm of up to -290 p.p.m. are clearly dominated by a proton-rich

process (p-process) deficiency. The lack of correlation between ^{142}Nd and ^{144}Sm variations between Solar System objects characterized by different ^{144}Sm signatures indicates that the contribution of p-processes to ^{142}Nd is negligible, contrary to previous suggestions^{4,21} (Extended Data Fig. 5). The positive $\mu^{148}\text{Sm}$ – $\mu^{150}\text{Sm}$ correlation (Extended Data Fig. 3) shows that CAIs are the carrier of a slow-neutron-capture-process

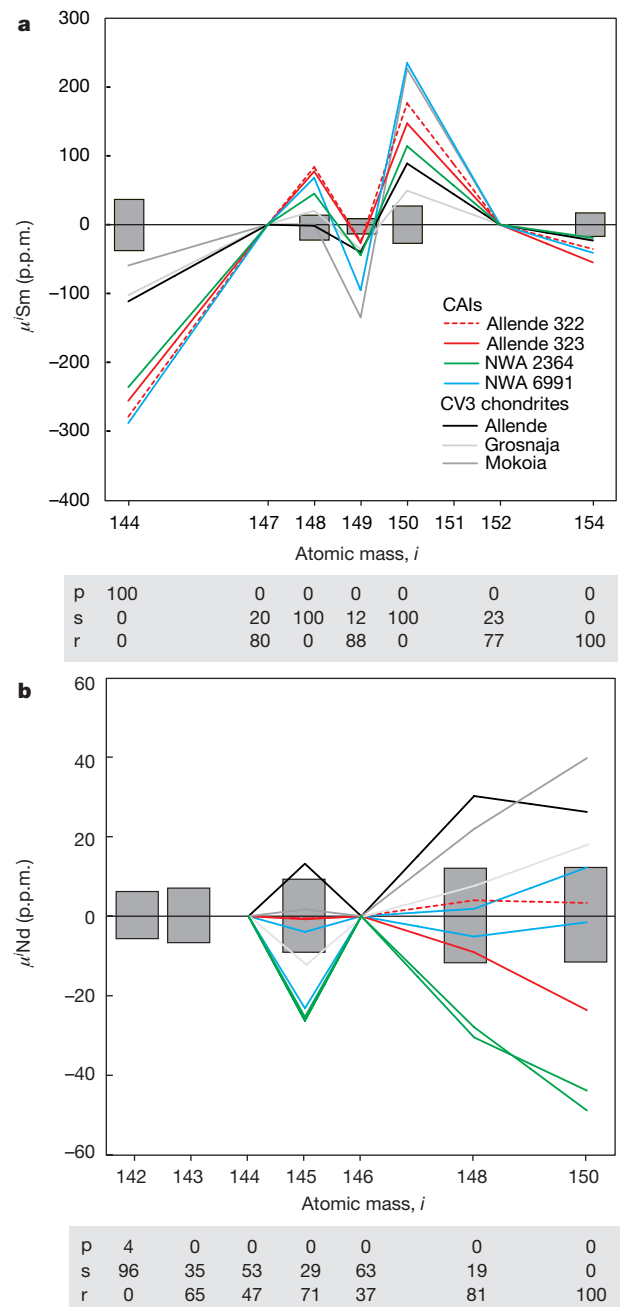


Figure 2 | Sm and Nd isotope compositions of bulk CAIs (leached and unleached) and CV3 whole-rocks. **a**, **b**, $\mu^{148}\text{Sm}$ (**a**) and $\mu^{142}\text{Nd}$ (**b**) represent the isotopic composition measured in meteorites relative to the isotope ratio measured in the terrestrial standard, and is given in parts per million. Sm and Nd data were corrected for instrumental mass fractionation using the exponential law and $^{147}\text{Sm}/^{152}\text{Sm} = 0.56081$ and $^{146}\text{Nd}/^{144}\text{Nd} = 0.7219$, respectively (Supplementary Tables 1–3). Grey boxes show the external reproducibility (2 s.d.) obtained on the standards. The relative contributions of the p-, s- and r-processes for stable and radiogenic (^{142}Nd and ^{143}Nd) isotopes are indicated (as percentages) below the plots²². The now-extinct ^{146}Sm is not represented, but is a pure p-process isotope. The two radiogenic Nd isotopes, ^{142}Nd and ^{143}Nd , are not represented because their deviations reflect mostly the radiogenic contribution.

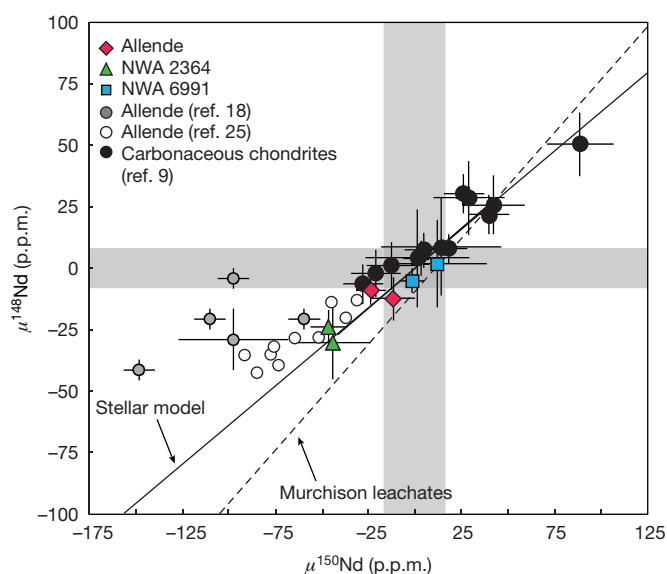


Figure 3 | $^{150}\text{Nd}/^{144}\text{Nd}$ versus $^{148}\text{Nd}/^{144}\text{Nd}$ measured in CAIs and carbonaceous chondrites. The data for carbonaceous chondrites are from ref. 9 and those for Allende CAIs are from refs 18 and 25. Error bars show the analytical uncertainty (2 s.e.); the grey boxes represent the external reproducibility (2 s.d.) obtained on the JNdi-1 standard. The internal errors for Allende bulk CAIs are not reported in ref. 25. The data plot on the mixing line calculated from isotope production in stellar models (solid black line)^{22,26}, but not on the regression defined by leachate Nd isotopic data of the CM2 chondrite Murchison (dashed black line)²⁴. The signatures in ^{148}Nd and ^{150}Nd of the Allende 322 and 323 and NWA 6991 CAIs are not different from the terrestrial standard considering the analytical uncertainties.

(s-process) excess and a rapid-neutron-capture-process (r-process) deficit in CV chondrites. The stable Nd isotope compositions of the bulk CAIs and their mineral fractions are similar to the terrestrial values within analytical error, except for the NWA 2364 CAI, which has deficits in ^{145}Nd , ^{148}Nd and ^{150}Nd (Fig. 2b). We identified a CAI in CV3 NWA 6991 that is petrogenetically primitive (Methods) and shares an identical Nd stable isotopic composition with Earth. Because different mineral fractions have been separated, the radiogenic ^{142}Nd isotope of NWA 6991 can be compared to the value for Earth without the need for any correction for neutron capture or nucleosynthetic anomalies. Its $\mu^{142}\text{Nd}$ value at a chondritic Sm/Nd ratio is -8.4 ± 5.1 p.p.m., which is indistinguishable within errors from the terrestrial value (0 ± 5 p.p.m.). By contrast, the s-process excesses and r-process deficits identified in the Nd stable isotopes of NWA 2364 CAIs should substantially increase the $^{142}\text{Nd}/^{144}\text{Nd}$ ratios (Fig. 2a, b). The correction for ^{142}Nd is complex and its accuracy is difficult to evaluate. Isotopes from the same element should be used to correct for nucleosynthetic effects³. A correction of -25 p.p.m. on ^{142}Nd is calculated using different approaches: (i) by using the ^{142}Nd – ^{148}Nd relationship calculated from a stellar model ($\mu^{148}\text{Nd} = -1.01 \times \mu^{142}\text{Nd}$; ref. 22); (ii) by using the isotope composition measured in SiC as representative of a pure s-process component ($\mu^{148}\text{Nd} = -1.15 \times \mu^{142}\text{Nd}$; ref. 23); and (iii) by considering the correlation in leachates data obtained on carbonaceous chondrites ($\mu^{148}\text{Nd} = -0.96 \times \mu^{142}\text{Nd}$; ref. 24). A smaller correction of 14 p.p.m. to the $^{142}\text{Nd}/^{144}\text{Nd}$ ratio was calculated for the Allende Al3S4 CAI¹⁸, whereas the separated mineral fractions of Al3S4 were characterized by the highest deficits in both ^{148}Nd and ^{150}Nd measured in CAIs. We note that the mineral separates from Allende Al3S4 CAI have different $\mu^{148}\text{Nd}$ and $\mu^{150}\text{Nd}$ values¹⁸ that do not fall on the correction line between CAIs and carbonaceous chondrites (Fig. 3). In contrast, carbonaceous chondrites and bulk CAIs from this study and from ref. 25 plot on the $\mu^{148}\text{Nd}$ – $\mu^{150}\text{Nd}$ mixing line calculated from isotope production in stellar models^{22,26} represented in Fig. 3. These CAIs do not fall on the regression line defined by Nd isotopic

data of leachates of the CM2 chondrite Murchison²⁴, which instead plot towards the composition of presolar SiC grains. Ultimately, after correcting for nucleosynthetic anomalies and radiogenic decay, CAIs show a range of $^{142}\text{Nd}/^{144}\text{Nd}$ ratios similar to enstatite chondrites, with values between the modern terrestrial average composition and those of ordinary chondrites. Furthermore, our data on enstatite chondrites have a range of $^{147}\text{Sm}/^{144}\text{Nd}$ ratios from 0.1948 to 0.1958, within 1% of the chondritic uniform reservoir (CHUR) value of 0.1960 (ref. 27), and do not show any correlation between $\mu^{142}\text{Nd}$ and $\mu^{144}\text{Sm}$ anomalies, in contrast to a previous study (Extended Data Fig. 5). We can therefore calculate the initial and modern compositions at a common chondritic $^{147}\text{Sm}/^{144}\text{Nd}$ without any model dependence and without introducing large errors on initial ^{142}Nd abundance. We calculate $\mu^{142}\text{Nd}$ values from -3 p.p.m. to -9 p.p.m. when normalized at the CHUR $^{147}\text{Sm}/^{144}\text{Nd}$ ratio (Supplementary Table 1). From our data, we obtain an average enstatite chondrite composition with $\mu^{142}\text{Nd} = -7 \pm 6$ p.p.m. (2. s.d.; $n = 3$), which is indistinguishable within errors from the composition of Earth's mantle of 0 ± 5 p.p.m. (Fig. 1). The dataset presented in ref. 4 had more scattered Sm/Nd ratios, which introduced uncertainties when correcting for ^{142}Nd produced by radiogenic decay and when normalizing $^{142}\text{Nd}/^{144}\text{Nd}$ ratios of individual whole-rock chondrites to a common CHUR Sm/Nd ratio. Nevertheless, we calculate a combined average for EL and EH chondrites (data from this study and ref. 4) of $\mu^{142}\text{Nd} = -9 \pm 17$ p.p.m. (2 s.d.; $n = 13$ individual enstatite chondrites), which is less precise, but remains consistent with Earth's composition.

We find that several early Solar System objects, including CAIs and enstatite chondrites, share a common Nd stable isotope signature with modern terrestrial samples, even though these materials sample spatially or temporally distinct regions of the protoplanetary disk. The CAIs formed under reduced conditions, closest to the proto-star, before being transported outward to the accretion regions of the chondrite groups²⁸. The reduced enstatite chondrites are suggested to have formed within the inner fringe of the asteroid belt, and the more oxidized ordinary chondrites and carbonaceous chondrites in the outer parts²⁹. Our results therefore suggest a relationship between ^{142}Nd abundance and heliocentric distance. Materials formed in the inner region of the Solar System have the highest ^{146}Sm decay-normalized $^{142}\text{Nd}/^{144}\text{Nd}$ ratios, followed by whole-rock ordinary chondrites and carbonaceous chondrites. The amount of material required to have been hidden within or lost from Earth was previously constrained using ^{142}Nd abundances measured in ordinary chondrites (see, for example, refs 5, 9). Although the building blocks of Earth are difficult to identify on the basis of meteorites that escaped accretion, enstatite chondrites are closest to Earth's composition when looking at the abundance of ^{142}Nd and of many other refractory elements¹⁴. Using the enstatite chondrite group as an isotope analogue of the accreting material that formed Earth, the calculated $^{147}\text{Sm}/^{144}\text{Nd}$ ratio of the silicate Earth would be 0.200, which is 2% higher than the CHUR $^{147}\text{Sm}/^{144}\text{Nd}$ ratio and within the range of values previously measured for whole-rock chondrites²⁷. An external constraint on the Sm/Nd ratio for Earth arises from the combined Sm–Nd and Lu–Hf systematics of lunar samples, suggesting a chondritic Sm/Nd evolution of the bulk Moon³⁰. Therefore, our results on CAIs together with those obtained on enstatite chondrites and lunar samples indicate that the Earth–Moon system evolved from a common isotopic reservoir in ^{142}Nd , with Sm/Nd proportions within 2% of the average of unequilibrated whole-rock chondrites²⁷. The increasing number of measurements of the ^{146}Sm – ^{142}Nd systematics of Solar System objects calls into question the notion of a super-chondritic Earth for refractory elements. Instead, our results indicate that variations in ^{142}Nd abundances found in planetary materials compared to the modern Earth are caused by nucleosynthetic anomalies, and eliminate the need for untapped or missing planetary reservoirs to explain the ^{146}Sm – ^{142}Nd isotope systematics of terrestrial rocks^{5,6}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 December 2015; accepted 4 August 2016.

- Nyquist, L. E. *et al.* ^{146}Sm - ^{142}Nd formation interval for the lunar mantle material. *Geochim. Cosmochim. Acta* **59**, 2817–2837 (1995).
- Boyett, M. & Carlson, R. W. ^{142}Nd evidence for early (>4.53 Ga) global differentiation of the silicate Earth. *Science* **309**, 576–581 (2005).
- Carlson, R. W., Boyett, M. & Horan, M. Chondrite barium, neodymium, and samarium isotopic heterogeneity and early Earth differentiation. *Science* **316**, 1175–1178 (2007).
- Gannoun, A., Boyett, M., Rizo, H. & El Goresy, A. ^{146}Sm - ^{142}Nd systematics measured in enstatite chondrites reveals a heterogeneous distribution of ^{142}Nd in the solar nebula. *Proc. Natl Acad. Sci.* **108**, 7693–7697 (2011).
- Jellinek, A. M. & Jackson, M. G. Connections between the bulk composition, geodynamics and habitability of Earth. *Nat. Geosci.* **8**, 587–593 (2015).
- O'Neill, H. S. C. & Palme, H. Collisional erosion and the non-chondritic composition of terrestrial planets. *Philos. Trans. R. Soc. Lond. A* **366**, 4205–4238 (2008).
- Kinoshita, N. *et al.* A shorter ^{146}Sm half-life measured and implications for ^{146}Sm - ^{142}Nd chronology in the Solar System. *Science* **335**, 1614–1617 (2012).
- Boyett, M., Carlson, R. W. & Horan, M. Old Sm-Nd ages for cumulate eucrites and redetermination of the Solar System initial $^{146}\text{Sm}/^{144}\text{Sm}$ ratio. *Earth Planet. Sci. Lett.* **291**, 172–181 (2010).
- Caro, G., Bourdon, B., Halliday, A. N. & Quitté, G. Super-chondritic Sm/Nd ratios in Mars, the Earth and the Moon. *Nature* **452**, 336–339 (2008).
- Clayton, R. N., Hinton, R. W. & Davis, A. M. Isotopic variations in the rock-forming elements in meteorites. *Philos. Trans. R. Soc. Lond. A* **325**, 483–501 (1988).
- Anders, E. & Zinner, E. K. Interstellar grains in primitive meteorites: diamond, silicon carbide, and graphite. *Meteoritics* **28**, 490–514 (1993).
- Trinquier, A. *et al.* Origin of nucleosynthetic isotope heterogeneity in the solar protoplanetary disk. *Science* **324**, 374–376 (2009).
- Moynier, F. *et al.* Planetary-scale strontium isotopic heterogeneity and the age of volatile depletion of early Solar System materials. *Astrophys. J.* **758**, 45 (2012).
- Warren, P. H. Stable-isotopic anomalies and the accretionary assemblage of the Earth and Mars: a subordinate role for carbonaceous chondrites. *Earth Planet. Sci. Lett.* **311**, 93–100 (2011).
- Clayton, R. N., Mayeda, T. K. & Rubin, A. E. Oxygen isotopic compositions of enstatite chondrites and aubrites. *Lunar Planet. Sci. Conf. Proc.* **15**, C245–C249 (1984).
- Bouvier, A. & Wadhwa, M. The age of the Solar System redefined by the oldest Pb–Pb age of a meteoritic inclusion. *Nat. Geosci.* **3**, 637–641 (2010).
- Kita, N. T. *et al.* ^{26}Al - ^{26}Mg isotope systematics of the first solids in the early Solar System. *Meteorit. Planet. Sci.* **48**, 1383–1400 (2013).
- Marks, N. E. *et al.* Samarium–neodymium chronology and rubidium–strontium systematics of an Allende calcium–aluminum-rich inclusion with implications for ^{146}Sm half-life. *Earth Planet. Sci. Lett.* **405**, 15–24 (2014).
- Meeker, G. P., Wasserburg, G. J. & Armstrong, J. T. Replacement textures in CAI and implications regarding planetary metamorphism. *Geochim. Cosmochim. Acta* **47**, 707–721 (1983).
- Gannoun, A., Boyett, M., El Goresy, A. & Devouard, B. REE and actinide microdistribution in Sahara 97072 and ALHA77295 EH3 chondrites: a combined cosmochemical and petrologic investigation. *Geochim. Cosmochim. Acta* **75**, 3269–3289 (2011).
- Andreasen, R. & Sharma, M. Solar nebula heterogeneity in p-process samarium and neodymium isotopes. *Science* **314**, 806–809 (2006).
- Arlandini, C. *et al.* Neutron capture in low-mass asymptotic giant branch stars: cross sections and abundance signatures. *Astrophys. J.* **525**, 886–900 (1999).
- Hoppe, P. & Ott, U. Mainstream silicon carbide grains from meteorites. *AIP Conf. Proc.* **402**, 27–58 (1997).
- Qin, L., Carlson, R. W. & Alexander, C. M. O. D. Correlated nucleosynthetic isotopic variability in Cr, Sr, Ba, Sm, Nd and Hf in Murchison and QUE 97008. *Geochim. Cosmochim. Acta* **75**, 7806–7828 (2011).
- Brennecka, G. A., Borg, L. E. & Wadhwa, M. Evidence for supernova injection into the solar nebula and the decoupling of r-process nucleosynthesis. *Proc. Natl Acad. Sci. USA* **110**, 17241–17246 (2013).
- Bisterzo, S. *et al.* Galactic chemical evolution and solar s-process abundances: dependence on the ^{13}C -pocket structure. *Astrophys. J.* **787**, 10 (2014).
- Bouvier, A., Vervoort, J. D. & Patchett, P. J. The Lu–Hf and Sm–Nd isotopic composition of CHUR: constraints from unequilibrated chondrites and implications for the bulk composition of terrestrial planets. *Earth Planet. Sci. Lett.* **273**, 48–57 (2008).
- Simon, J. I. *et al.* A short timescale for changing oxygen fugacity in the solar nebula revealed by high-resolution ^{26}Al - ^{26}Mg dating of CAI rims. *Earth Planet. Sci. Lett.* **238**, 272–283 (2005).
- Jacquet, E. Transport of solids in protoplanetary disks: comparing meteorites and astrophysical models. *C. R. Geosci.* **346**, 3–12 (2014).
- Sprung, P., Kleine, T. & Scherer, E. E. Isotopic evidence for chondritic Lu/Hf and Sm/Nd of the Moon. *Earth Planet. Sci. Lett.* **380**, 77–87 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements Meteorite samples were provided by L. Garvie (Arizona State University), D. Ebel (American Museum of Natural History) and the US Antarctic Search for Meteorites (ANSMET) programme, which has been funded by NSF and NASA, and were characterized and curated by the Department of Mineral Sciences of the Smithsonian Institution and Astromaterials Curation Office at NASA Johnson Space Center. We thank P. J. Patchett for the Sm–Nd calibrated enriched spike used in this study, D. Auclair and A. Gannoun for mass spectrometer support, and T. Withers for comments on this Letter. This research was supported by the National Science Foundation NSF/EAR 1119135, France-Canada Research Fund, NSERC Canada Research Chair and Discovery Grant awards to A.B. and the French Government ANR-10-LABX-0006, the Région Auvergne, the European Regional Development Fund and the INSU Programme National de Planétologie to M.B. We thank the Laboratory of Excellence ClerVolc.

Author Contributions A.B. and M.B. planned and carried out the analyses for the project and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.B. (audrey.bouvier@uwo.ca).

METHODS

Samples and analytical methods. Calcium–aluminium-rich inclusions (CAIs) were obtained from the Center for Meteorite Studies at Arizona State University for Allende (named 322 and 323) and Northwest Africa (NWA) 6991 (named B4, fragment mass of about 150 mg), and from the American Museum of Natural History in New York for NWA 2364 (named the ‘crucible’, fragment mass of about 150 mg). For Allende, the two small (50–100 mg), fine-grained (unmelted) CAIs were processed as bulk powders. These four CAIs were selected on the basis of their petrogenetic differences, to assess potential Sm and Nd stable isotopic heterogeneities linked to different CAI groups and formation histories. Geochemical studies of fine-grained CAIs in CV3 chondrites indicate that they did not melt and that they belong to group II CAIs, on the basis of their rare-earth elemental abundances³¹. These characteristics point towards formation as a gas–solid condensate from the nebular gas with minor reprocessing³¹. The Sm and Nd concentrations normalized to CI chondrites are $Sm_N = 19–20$ and $Nd_N = 20–22$. The lower, normalized Lu concentrations of $Lu_N = 0.3–3$ (concentrations in Extended Data Table 1) that were measured in these two CAIs by isotopic dilution point towards a fractionated rare-earth element (REE) pattern and a group II classification for both Allende CAIs 322 and 323 (Extended Data Fig. 1). The inclusion B4 of NWA 6991 was petrographically characterized² as a compact type A CAI. The CAIs B4 and crucible were processed for mineral separation. Mineral separates (melilite and fassaite) were concentrated from a 30–63- μ m fraction using bromoform and methylene iodide heavy liquids, and hand picking as described in ref. 16. The same CAI B4 has been dated using the Al–Mg and U–Pb radiogenic decay systems at $4,567.94 \pm 0.31$ Myr (ref. 32). Bulk samples were processed at Arizona State University (Allende) and the University of Minnesota (NWA 6991 and NWA 2364) as unleached and as leached with 10% HF and, subsequently, with 20% HCl for 20 min each at room temperature. Mineral separates were similarly leached before dissolution. The leachates were not analysed for ^{146}Sm – ^{142}Nd systematics owing to the limited leached sample masses.

Enstatite chondrites were obtained from the US Antarctic meteorite collection. Chips of about 1.2–1.4 g of enstatite chondrites MacAlpine Hills (MAC) 02837, 26, MAC 02839, 12 and Allan Hills A77295 (ALHA77295 or ALH 77295), 74 were crushed in an agate mortar and pestle into fine whole-rock powders by separating and remixing the metal to preserve all components together before acid digestion.

New PFA Savillex beakers were used for these samples to avoid memory effects. Parr bombs with 14-ml insert PTFE vials were not previously exposed to enriched spikes and were used for only meteorite dissolution. Only MilliQ Millipore water, BDH Aristar Ultra hydrofluoric and hydrochloric acids, and triple sub-boiled quartz-distilled nitric acid were used for all steps.

For enstatite chondrites, the whole-rock powders were dissolved using a first step in concentrated HF–HNO₃ (10:1) at 120 °C on a hot plate for 2 days to liberate hydrogen sulfide and silicate tetrafluoride gases before transferring the residues into Parr pressured vials. For CAIs and enstatite chondrites, samples were placed into individual Parr bomb pressured vials in concentrated HF–HNO₃ (10:1) for 7 days at 155 °C. After drying down the samples, perchloric acid was added and evaporated at 200 °C on a hot plate to break down the fluorides, and samples were taken back into solution in Parr bomb pressured vials in 6 M HCl for 2 days at 155 °C. Solutions were transferred into Savillex PFA beakers and split into an approximately 90% fraction for isotopic composition, an approximately 5%–10% fraction spiked with an enriched ^{149}Sm – ^{150}Nd mixed spike for isotopic dilution to determine the Sm/Nd ratio (using the measured stable Sm and Nd isotopic compositions), and an approximately 5% fraction for inductively coupled plasma mass spectrometry (ICPMS) analysis for trace element abundances for the NWA 6991 CAI bulk sample (Extended Data Fig. 1). The CAI B4 has a flat REE pattern normalized to CI chondrites ($(17–20) \times$ CI chondrites) with slight positive Eu anomaly ($Eu_N = 23$) corresponding to a chemical group I CAI⁷. The same group I is suggested for NWA 2364 unleached bulk CAI with $Sm_N = 17$, $Nd_N = 16$ and $Lu_N = 20$ (CI-chondrite normalized REE concentrations shown in Extended Data Fig. 1 and concentrations given in Extended Data Table 1).

All Sm and Nd concentrations obtained by isotope dilution are reported in Extended Data Table 1. Sm and Nd in the spiked fractions were purified and separated using a two-stage chemistry procedure. The REEs were first separated from the matrix using a 2-ml or 8-ml cation resin AG50W-X8 column. The REE fraction was then loaded in 0.14 M HCl onto a 1-ml Eichrom Ln-spec resin bed to separate the Nd and Sm fractions in 0.14 M HCl and 0.40 M HCl, respectively. The chemistry procedure was different for the unspiked fractions. After the first 8-ml cation columns, the REE fractions were processed twice through a cation column using 2-methylactic acid (0.2 M, pH 4.7) with a small amount of H₂O₂. The Sm fraction was collected before the Nd fraction during the first chemistry, but at this stage heavier REEs were still present within the two collected fractions. The last step consists of one pass through Ln-spec resin in weak HCl. This

procedure ensures perfect separation of Nd from Ce and Sm by reducing the effect of interferences on mass 142 from Ce and on masses 144, 148 and 150 from Sm, and the purification of the Sm fraction (Gd interferes at masses 152 and 154). Organic residues are then completely removed using hydrogen peroxide and the sample is ready to be loaded on a Re filament. Total procedural blanks for Sm and Nd were 3 pg and 10 pg, respectively. The blank contribution to each sample was negligible, with Sm and Nd fractions of at least 80 ng for the smallest analysed sample.

Spiked Sm and Nd isotopic fractions for the isotopic dilution method were analysed in static mode using a MC-ICPMS Neptune at Arizona State University for Allende CAIs, and Neptune Plus at the University of Minnesota and the Laboratoire Magmas et Volcans (Clermont-Ferrand, France) for the NWA 6991 and NWA 2364 CAIs using methods described in ref. 27. AMES Sm and JNdi-1 Nd standards were used to control mass bias and cup efficiency during all the sessions. The calculated Sm and Nd concentrations and $^{147}Sm/^{144}Nd$ ratio of samples are reported in Supplementary Table 1. Additionally, a BCR-2 rock standard was processed and analysed in each laboratory, and Sm and Nd concentrations and calculated $^{147}Sm/^{144}Nd$ and $^{143}Nd/^{144}Nd$ ratios are reported in Supplementary Table 1.

Sm and Nd isotopes of unspiked fractions were obtained on the Thermo-Fisher Triton thermal ionization mass spectrometer at the Laboratoire Magmas et Volcans. The purified Sm and Nd cuts were loaded in 2.5 M HCl on outgassed, zone-refined Re filaments. The Sm was measured in static mode as Sm^+ . The Faraday cup configuration was centred at mass 149 and Nd and Gd interferences were monitored at masses 146 (cup L3) and 156 (cup H4). Each run consisted of 18 to 27 blocks of 20 cycles using amplifier rotation with background acquisition before each block. Sample measurements were in general shorter depending on the amount loaded (see Supplementary Table 2). Signals for ^{152}Sm ranged between 0.5×10^{-11} A and 3.5×10^{-11} A for standards and between 0.5×10^{-11} A and 2.0×10^{-11} A for samples. Sm data were corrected for instrumental mass fractionation using the exponential law and $^{147}Sm/^{152}Sm = 0.56081$. Sm isotope compositions for standards and samples are given in Supplementary Table 2. Sm isotope compositions were measured during three different periods, called session #1 to #3 in Supplementary Table 2. The μ values correspond to deviations relative to the standard value, expressed in parts per million. In static mode, the ratios evolve with time owing to cup ageing. Consequently, we used the average of ratios measured on standards during the same analytical session for calculating the sample μ values. The external reproducibility calculated from repeated measurements of the standard (5–7 standards measured per analytical session) is 15–37 p.p.m. on $^{144}Sm/^{152}Sm$, 5–18 p.p.m. on $^{148}Sm/^{152}Sm$, 7–11 p.p.m. on $^{149}Sm/^{152}Sm$, 14–27 p.p.m. on $^{150}Sm/^{152}Sm$ and 12–17 p.p.m. on $^{154}Sm/^{152}Sm$. The Nd samples were measured as Nd^+ using 9 Faraday cups. We used a dynamic procedure (axial cup centred at masses 143 and 145) to measure the $^{142}Nd/^{144}Nd$ ratios in dynamic mode; all other ratios were measured in static mode (line 1 centred at 145). Ce and Sm interferences on masses 142 and 144 were monitored using masses 140 and 147, respectively. Ce and Sm contributions on masses 142 and 144 were lower than 3 p.p.m. for standard and sample measurements, except for sample NWA 6991, in which the Sm contribution on mass 144 reaches 15 p.p.m. Nd isotope ratios were corrected for mass fractionation to $^{146}Nd/^{144}Nd = 0.7219$ using the exponential law. The external reproducibility (2 s.d.) calculated from repeated measurements of the standard within each session is 4–6 p.p.m. on $^{142}Nd/^{144}Nd$, 3–7 p.p.m. on $^{143}Nd/^{144}Nd$, 4–8 p.p.m. on $^{145}Nd/^{144}Nd$, 3–12 p.p.m. on $^{148}Nd/^{144}Nd$ and 10–12 p.p.m. on $^{150}Nd/^{144}Nd$ (Supplementary Table 3).

Sm isotopes. The Sm isotopic compositions measured for terrestrial standards and samples are presented in Supplementary Table 2. For samples measured during the first sequence, the quantity of Nd is substantially higher than that present in the Sm standard. Nevertheless, there is no correlation between Sm isotope ratios and the quantity of Nd found in our samples (Extended Data Fig. 2). The ^{144}Sm deficits in all bulk CAI samples are in the range –236 p.p.m. to –288 p.p.m. and do not correlate with Nd/Sm (Extended Data Fig. 2). The melilite fraction of the NWA 2364 CAI has the lowest deficit in ^{144}Sm ($\mu^{144}Sm = -113$ p.p.m.) and lower Sm isotope anomalies than other CAI fractions in general. This sample was measured with the lowest Sm intensity and during a very short period (60 ratios), so isotopic ratios have a precision that is three times lower than those of the other samples. All CAI fractions, except for the NWA 6991 leached bulk fraction and the melilite from NWA 2364, have excesses in ^{148}Sm (+45 p.p.m. to +84 p.p.m.), deficits in ^{149}Sm (–25 p.p.m. to –95 p.p.m.), excesses in ^{150}Sm (+114 p.p.m. to +240 p.p.m.) and deficits in ^{154}Sm (–18 p.p.m. to –54 p.p.m.). The Sm isotope patterns are presented in Fig. 2a. In this representation, the Sm isotopes at masses 147 and 152 are equal to zero because measured data are normalized using the $^{147}Sm/^{152}Sm$ ratio.

Deficits in ^{144}Sm can reflect deficits in p-process isotopes, but they are also artificially created by s-process excess and/or r-process deficits using the $^{147}Sm/^{152}Sm$ normalization scheme. The two Sm isotopes used for mass bias correction are predominantly formed by r-processes (88% and 77% for ^{147}Sm and ^{152}Sm , respectively), and the remaining part is formed by s-processes. The positive

$\mu^{148}\text{Sm}-\mu^{150}\text{Sm}$ correlation (Extended Data Fig. 3) shows that CAIs are the carrier of an s-process excess and an r-process deficit. This correlation is in agreement with the general shape of the pattern, with deficits in ^{144}Sm , ^{149}Sm and ^{154}Sm , and excesses in ^{148}Sm and ^{150}Sm (Fig. 2a). The NWA 6991 CAI has the highest deficits in ^{149}Sm and highest excesses in ^{150}Sm , and this sample does not fall on the $\mu^{148}\text{Sm}-\mu^{150}\text{Sm}$ correlation line formed by CV3 chondrites and other CAIs²⁵. This result suggests that the Sm isotope composition of the NWA 6991 CAI has been modified by neutron fluence capture, owing to long-term exposure to galactic cosmic rays following the reaction $^{149}\text{Sm}(n, \gamma)^{150}\text{Sm}$. Nevertheless, this effect is small and does not induce any substantial modification of Nd isotope ratios because Nd isotopes have smaller neutron capture cross sections than does ^{149}Sm (ref. 1). CAIs and carbonaceous chondrites plot on the same correlation line, which intersects the ^{148}Sm terrestrial value at $\mu^{144}\text{Sm} = -60$ p.p.m. in the $\mu^{144}\text{Sm}-\mu^{148}\text{Sm}$ diagram (Extended Data Fig. 3). This value is in agreement with the deficit in ^{144}Sm of about -80 p.p.m. measured in carbonaceous chondrites that have terrestrial ^{148}Sm abundances^{3,21}. The CAIs represent the end-member farthest from the terrestrial composition. The three enstatite chondrites have terrestrial compositions for both ^{148}Sm and ^{154}Sm , indicating that the $^{149}\text{Sm}-^{150}\text{Sm}$ signature (excess in ^{149}Sm coupled to deficit in ^{150}Sm) reflects secondary reactions. Deficit or excess in s-processes must generate similar deviations in both $^{148}\text{Sm}/^{152}\text{Sm}$ and $^{150}\text{Sm}/^{152}\text{Sm}$ ratios. The two EL3 samples MAC 02837 and 02839 have variable ^{144}Sm abundances with μ values of -3 ± 16 p.p.m. and 48 ± 12 p.p.m., respectively.

Nd isotopes. The Nd isotopic compositions measured for CAIs and enstatite chondrite samples are presented in Supplementary Table 3. $^{142}\text{Nd}/^{144}\text{Nd}$ ratios were measured in static and dynamic modes. Data are compared in Extended Data Fig. 4 and ratios fall on the 1:1 line. The contribution of Ce and Sm measured in JNdi-1 standard and samples is negligible, with the exception of the NWA 2364 CAI measured during the third session, in which the cerium correction on the $^{142}\text{Nd}/^{144}\text{Nd}$ ratio is 560 p.p.m. The stable Nd isotope compositions of the bulk CAIs and their mineral fractions are similar to the terrestrial values within the analytical error, except for the NWA 2364 CAI, which has deficits in ^{145}Nd ($\mu = -26 \pm 4$ p.p.m.), ^{148}Nd ($\mu = -25 \pm 6$ p.p.m.) and ^{150}Nd ($\mu = -48 \pm 9$ p.p.m.) (Fig. 2b). As observed for Sm isotopes, this signature results from s-process excesses and r-process deficits. The Nd isotope patterns are presented in Fig. 2b. Using this representation, the Nd at masses 144 and 146 are equal to zero, because the measured data are normalized using the $^{146}\text{Nd}/^{144}\text{Nd}$ ratio. For Nd isotopes in enstatite chondrites, MAC 02837 has a terrestrial abundance in ^{142}Nd ($\mu = -6 \pm 7$), whereas MAC 02839 and ALHA77295 have small deficits ($\mu = -10 \pm 6$ and $\mu = -11 \pm 3$, respectively). For stable isotopes, the abundances are identical to those measured in Nd standards and terrestrial samples, except for MAC 02837, which has a deficit of -115 ± 9 p.p.m. in ^{150}Nd abundance. Because all other Nd isotopes are present in terrestrial abundances, the cause of this deficit in ^{150}Nd is unclear.

Neutron fluence effect on the Sm isotope patterns of CAIs. The Sm isotope patterns for CAIs have deficits in ^{149}Sm and excesses in ^{150}Sm . The negative correlation between the $^{149}\text{Sm}/^{152}\text{Sm}$ and $^{150}\text{Sm}/^{152}\text{Sm}$ ratios (Extended Data Fig. 3) may reflect deficits or excesses in s- and r-process nuclides because ^{149}Sm is dominated by r-processes (88%), whereas ^{150}Sm is a pure s-process nuclide. This correlation can also be created by nuclear reactions induced by interaction between galactic cosmic rays and the surface of extraterrestrial material. At low energy, epithermal and thermal neutron reactions produce nuclides by neutron capture. ^{149}Sm has the largest neutron capture cross-section of all Sm isotopes and the reaction $^{149}\text{Sm}(n, \gamma)^{150}\text{Sm}$ is commonly used for estimating the amplitude of this secondary nuclear reaction in lunar samples (black line in Extended Data Fig. 3a drawn from lunar data³³). Nuclear reactions induced by exposure to galactic cosmic rays are the result of a surface process that becomes negligible at relatively small depths (below 2–3 m; ref. 34). The deficits in ^{149}Sm measured in Allende CAIs (-25 p.p.m. to -31 p.p.m.) are similar to those measured in the different dissolutions of the Allende bulk sample^{3,21}, but the excesses in ^{150}Sm in CAIs ($+147$ p.p.m. to $+176$ p.p.m.) are higher than those measured in the whole rocks ($+90$ p.p.m.). The Sm isotope patterns of Allende CAIs are not affected by neutron fluence. The NWA 6991 CAI has both a higher deficit in ^{149}Sm (-95 p.p.m.) and a higher excess in ^{150}Sm ($+235$ p.p.m.) than other CAIs measured in this study, and plots close to the whole-rock CV3 Mokoia in Fig. 2a (data from ref. 3). The NWA 6991 CAI also falls far from the regression line for Allende, NWA 2364 and the CV3 whole-rock chondrites in the $\mu^{148}\text{Sm}-\mu^{150}\text{Sm}$ diagram (except Mokoia, Extended Data Fig. 3b). Because ^{148}Sm and ^{150}Sm are pure s-process nuclides, the two $^{148}\text{Sm}/^{152}\text{Sm}$ and $^{150}\text{Sm}/^{152}\text{Sm}$ ratios should be positively correlated for samples that have not been affected by interaction with galactic cosmic rays. The excess observed in ^{150}Sm for the NWA 6991 CAI is explained by secondary production of this isotope, and, using the correlation plotted in Extended Data Fig. 3b, we can estimate an increase in $\mu^{150}\text{Sm}$ of about 100 p.p.m. Using the $^{149}\text{Sm}-^{150}\text{Sm}$ relationships established for lunar samples³³, we calculate that an excess of about 100 p.p.m. in ^{150}Sm should correspond to a decrease in $\mu^{149}\text{Sm}$ of about 40 p.p.m. Such a variation will have

a negligible effect (<1 p.p.m.) on the corrected $^{142}\text{Nd}/^{144}\text{Nd}$ ratio, considering the relationship^{1,35} between Sm and Nd isotopes.

P-process contribution to ^{142}Nd production. ^{142}Nd is not a pure s-process isotope, but contains a small fraction of p-process isotope that varies between 4% and 20%, with the lower estimate obtained from stellar nucleosynthesis models³⁶ and the upper estimate from measurements of meteorites⁴. By monitoring the abundance of the pure p-process ^{144}Sm isotope in CAIs, we evaluate how this process affects the corresponding $^{142}\text{Nd}/^{144}\text{Nd}$ ratio. The Sm isotope patterns of all CAIs are similar to those measured in CV3 chondrites, but the anomalies (whether excess or deficit) are commonly larger (Fig. 2a), with two major differences. First, CAIs have large deficits in ^{144}Sm , up to -290 p.p.m. relative to the bulk CV (-100 p.p.m. on average^{3,21}). Second, they have small excesses in ^{148}Sm , whereas CV3 chondrites have terrestrial ^{148}Sm abundances. Deficits in ^{144}Sm can reflect deficits in p-process isotopes, but apparent deficits may also arise from the $^{147}\text{Sm}/^{152}\text{Sm}$ normalization used to correct instrumental mass bias when applied to samples with s-process excesses and/or r-process deficits. The contribution of the deficits induced by the normalization is estimated using stellar models^{22,26}, and shows that the ^{144}Sm deficit should be about 25% of the excess in ^{148}Sm , which corresponds to a maximum deficit of 15 p.p.m. in ^{144}Sm . The measured deficits in $\mu^{144}\text{Sm}$ of >200 p.p.m. in CAIs are therefore clearly dominated by a p-process deficit.

It has been shown³⁶ that a fractionated Nd/Sm p-process isotope contribution ratio resulting from chemical separation upon condensation in a circumstellar environment could also explain the $\mu^{144}\text{Sm}-\mu^{142}\text{Nd}$ isotope correlation found in ref. 4, with as little as 4% p-process contribution on ^{142}Nd . Combined Sm–Nd isotope measurements of different Solar System objects show that materials with a large range of $\mu^{144}\text{Sm}$ values have identical $\mu^{142}\text{Nd}$ compositions, as illustrated for the measured CAIs and the FUN (fractionated and unknown nuclear effects) inclusion C1 in Extended Data Fig. 5. The CAIs and FUN C1 fall on the 1% p-process–99% s-process ^{142}Nd line, which suggests that the contribution of p-process to ^{142}Nd is negligible and within our measurement errors. Moreover, after including our enstatite chondrite data, the $\mu^{144}\text{Sm}-\mu^{142}\text{Nd}$ correlation that had been previously found between the different groups of chondrites⁴ is no longer present. Only the carbonaceous chondrites are clearly distinct, with the highest deficits in both ^{144}Sm and ^{142}Nd . All the different mineral fractions and the bulk CAIs show excesses in s-process isotopes and deficits in r-process isotopes for Sm, whereas deviations in Nd isotope ratios relative to the terrestrial standard are observed for only NWA 2364 (Fig. 2b). Our data also confirm the variations in r-process deficits with a stronger effect on Sm in comparison to Nd (ref. 25).

The abundance of p-process isotopes relative to mass is shown in Extended Data Fig. 6. For isotopes formed by both p- and s-processes, the contribution of p-process isotopes are calculated using data from ref. 22 as the solar abundance minus the contribution of the s-process isotopes. The abundance of p-processes (normalized to Si) was recalculated using the total abundance of isotopes given in ref. 37. Extended Data Fig. 6 shows that the relation is well approximated by a power law. The abundance of pure p-process nuclides falls close to the curve. However, when we compare abundances calculated for ‘mixed’ isotopes (formed by mixing of p- and s-process isotopes such as ^{76}Se , ^{80}Kr , ^{142}Nd , ^{152}Gd and ^{164}Er), the abundances fall far from the calculated curve. If the ^{142}Nd was formed by 1% of p-process isotopes, it should fall close to the curve.

$^{146}\text{Sm}-^{142}\text{Nd}$ isochrons. The $^{147}\text{Sm}-^{143}\text{Nd}$ isotopic compositions of the CAIs are shown in Extended Data Fig. 7a. We obtain an absolute $^{147}\text{Sm}-^{143}\text{Nd}$ age of $4,519 \pm 140$ Myr (MSWD = 0.77, initial $^{143}\text{Nd}/^{144}\text{Nd} = 0.50675 \pm 0.00020$) for all CAI fractions using a half-life value for ^{147}Sm of $(6,539 \pm 0.061) \times 10^{-12}$ yr⁻¹ (ref. 38). The initial $^{143}\text{Nd}/^{144}\text{Nd}$ is in agreement with previous determinations of the initial composition of $^{143}\text{Nd}/^{144}\text{Nd}$ in the Solar System of 0.50669 ± 0.00007 (ref. 27). If we calculate individual internal CAI isochrons, we find the most precise internal age for NWA 6991 of $4,523 \pm 150$ Myr (MSWD = 1.7, initial $^{143}\text{Nd}/^{144}\text{Nd} = 0.50673 \pm 0.00021$). Our $^{147}\text{Sm}-^{143}\text{Nd}$ isochron ages are less precise than the Allende CAI age of $4,560 \pm 34$ Myr (ref. 18). A larger dispersion may come from the fact that we include four different CAIs from three different meteorites with potentially various evolution and metamorphic histories. Our mineral fractions have a limited range of Sm/Nd ratios in comparison to data presented in ref. 18. The melilite of NWA 2364 and melilite-residue of NWA 6991 with the lowest Sm/Nd ratio were too small for precise $^{142}\text{Nd}/^{144}\text{Nd}$ analysis and thus spiked for $^{143}\text{Nd}/^{144}\text{Nd}$ analysis.

In the $^{144}\text{Sm}/^{144}\text{Nd}$ versus $^{142}\text{Nd}/^{144}\text{Nd}$ diagram (Extended Data Fig. 7b), the slope of the best-fit line gives the $^{146}\text{Sm}/^{144}\text{Sm}$ ratio at the time of isotopic closure. When data obtained on the four CAIs from Allende, NWA 6991 and NWA 2364 are plotted together, they yield a $^{146}\text{Sm}/^{144}\text{Sm}$ ratio of 0.0073 ± 0.0022 (MSWD = 3.3) (Fig. 1). This ratio is indistinguishable (within errors) from the value of 0.0083 ± 0.0004 defined from the study of the Allende Al3S4 CAI¹⁸.

In conclusion, the initial ^{146}Sm abundance deduced from CAI isochrons is inconsistent with the $^{146}\text{Sm}/^{144}\text{Sm}$ ratio obtained from the initial eucrite isochron calculated back to the age of CAIs using a half-life of 68 Myr (ref. 7). Instead, our

result and those of ref. 18 support the use of a ^{146}Sm half-life of 103 Myr (ref. 39). Having a good estimate of the ^{146}Sm half-life is essential for using this short-lived chronometer. As shown in Extended Data Fig. 8, such variation in the ^{146}Sm half-life produces age differences of up to 150 Myr for Hadean rocks.

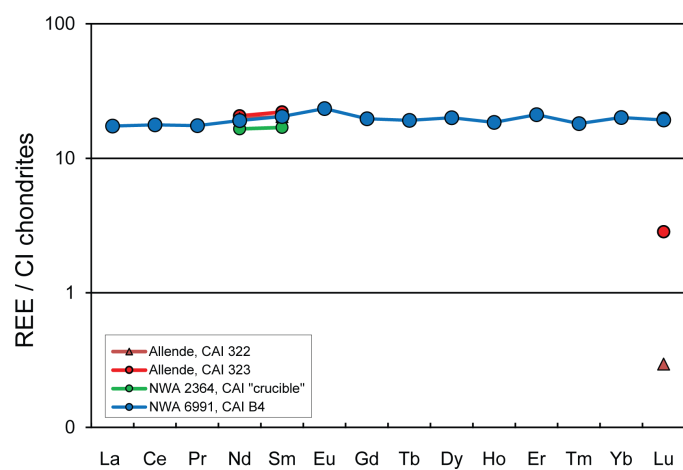
Influence of CAIs on the bulk whole-rock carbonaceous chondrite Sm isotope compositions. The abundance of CAIs varies among chondrite groups, and CV chondrites contain on average about 3 vol% CAIs⁴⁰. In CM chondrites, the proportion of CAIs is lower than in CV chondrites, and bulk samples are also characterized by lower deficits in ^{144}Sm (refs 3, 21). When we compare the measured ^{144}Sm negative anomalies measured in CAIs (this study) with enstatite, ordinary and carbonaceous whole-rock chondrites (this study and refs 3, 20) (Extended Data Fig. 9), we find that a volume abundance of CAIs in carbonaceous chondrites of approximately 3%–5% can explain the anomalies found in bulk CV3 (and CM2) chondrites, whereas enstatite chondrites and ordinary chondrites do not have any ^{144}Sm anomalies. However, the large deficit of ^{144}Sm in the Orgueil meteorite (CI chondrite, $\mu^{144}\text{Sm} = -103 \pm 46$ p.p.m., ref. 3) cannot be explained by the presence of CAIs because the abundance of CAIs in CI chondrites is lower than 0.01% (ref. 40). Therefore, ^{144}Sm deficits are inherent to the matrix of CI chondrites instead of CAIs in CV and CM chondrites. This suggests that there are p-process deficits in all carbonaceous chondrite groups, hosted within the matrix in CI chondrites or within CAIs in other carbonaceous chondrite groups.

Conversely, CAIs have s-process deficits and r-process excesses, whereas the bulk rock isotope compositions tend towards a component that has an opposite signature, and that could be similar to the FUN inclusion EK1-4-1 (refs 41, 42), which display large excesses in r-process isotopes (Extended Data Fig. 5). Deficits in ^{144}Sm are carried by CAIs in CV chondrites, whereas deficits in ^{142}Nd are carried within the matrix of chondrites, and are the most extreme in carbonaceous chondrites.

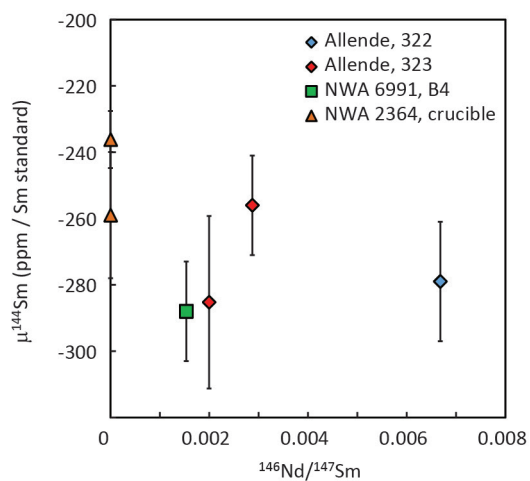
The CAIs measured in this study have $^{147}\text{Sm}/^{144}\text{Nd}$ ratios between 0.1755 and 0.2405. Adding 3%–5% CAIs would increase the $^{147}\text{Sm}/^{144}\text{Nd}$ ratio of the CV bulk chondrites to 0.198. Literature ^{147}Sm – ^{143}Nd data on Allende were reported in the supporting information of ref. 4. Measured Sm/Nd ratios for CV bulk rocks range from 0.1829 (ref. 35) to 0.2262 (ref. 43). These two extreme values were measured for very small Allende samples (90 mg and 0.6 mg, respectively). The dissolution

of Allende bulk-rock chips of several hundreds of milligrams to 1 g gives values closer to the average chondritic ratio of 0.1960 (ref. 8). Five CV3 whole-rock chondrites were previously measured²⁷ to have an average $^{147}\text{Sm}/^{144}\text{Nd}$ ratio of 0.1955 ± 0.0058 (2 s.d.).

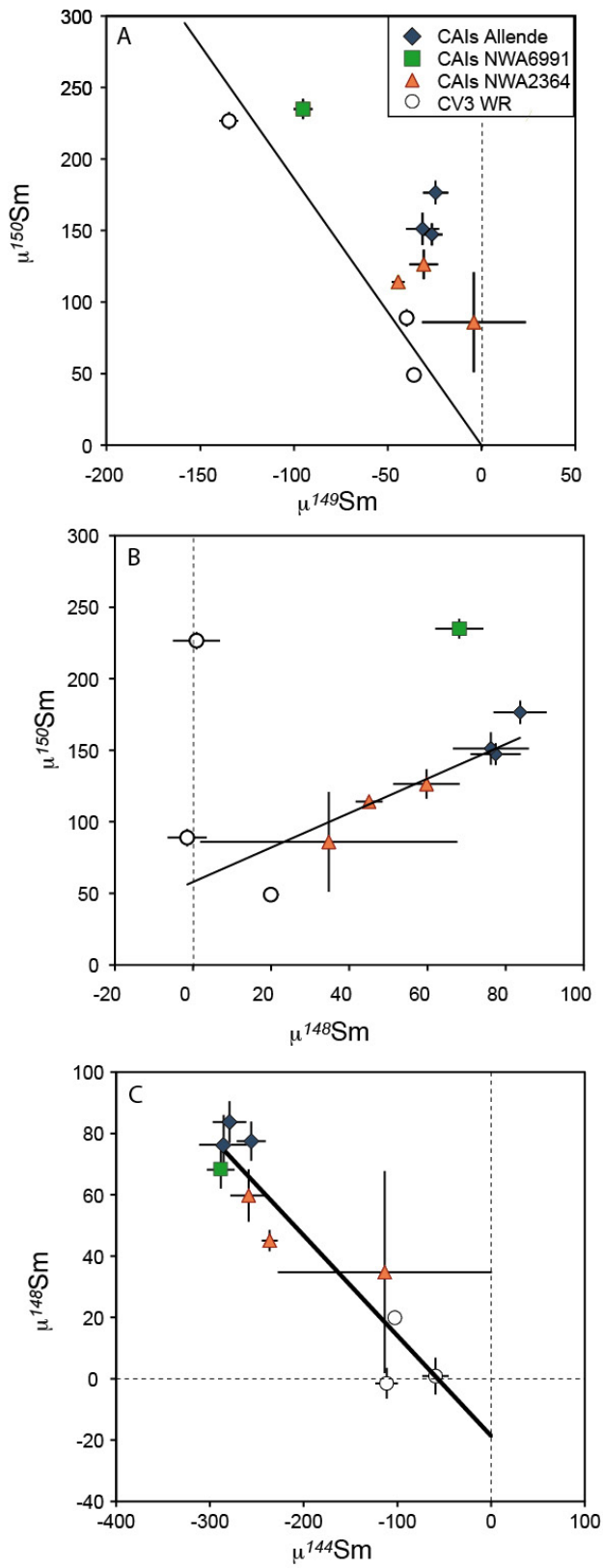
31. Krot, A. N., MacPherson, G. J., Ulyanov, A. A. & Petaev, M. I. Fine-grained, spinel-rich inclusions from the reduced CV chondrites Efremovka and Leoville: I. Mineralogy, petrology, and bulk chemistry. *Meteorit. Planet. Sci.* **39**, 1517–1553 (2004).
32. Bouvier, A., Brennecka, G. A. & Wadhwa, M. Absolute chronology of the first solids in the solar system. In *Formation of the First Solids in the Solar System* abstr. 9054 (2011).
33. Boyet, M. & Carlson, R. W. A highly depleted moon or a non-magma ocean origin for the lunar crust? *Earth Planet. Sci. Lett.* **262**, 505–516 (2007).
34. Marti, T. & Graf, T. Cosmic-ray exposure history of ordinary chondrites. *Annu. Rev. Earth Planet. Sci.* **20**, 221–243 (1992).
35. Rankenburg, K., Brandon, A. D. & Neal, C. R. Neodymium isotope evidence for a chondritic composition of the Moon. *Science* **312**, 1369–1372 (2006).
36. Rauscher, T. *et al.* Constraining the astrophysical origin of the p-nuclei through nuclear physics and meteoritic data. *Rep. Prog. Phys.* **76**, 066201 (2013).
37. Anders, E. & Grevesse, N. Abundances of the elements: meteoritic and solar. *Geochim. Cosmochim. Acta* **53**, 197–214 (1989).
38. Begemann, F. *et al.* Call for an improved set of decay constants for geochronological use. *Geochim. Cosmochim. Acta* **65**, 111–121 (2001).
39. Meissner, F., Schmidt-Ott, W. D. & Ziegeler, L. Half-life and α -ray energy of ^{146}Sm . *Z. Phys. A* **327**, 171–174 (1987).
40. Hezel, D. C., Russell, S. S., Ross, A. J. & Kearsley, A. T. Modal abundances of CAIs: implications for bulk chondrite element abundances and fractionations. *Meteorit. Planet. Sci.* **43**, 1879–1894 (2008).
41. McCulloch, M. T. & Wasserburg, G. J. Barium and neodymium isotopic anomalies in the Allende meteorite. *Astrophys. J.* **220**, L15–L19 (1978).
42. McCulloch, M. T. & Wasserburg, G. J. More anomalies from the Allende meteorite: samarium. *Geophys. Res. Lett.* **5**, 599–602 (1978).
43. Amelin, Y. & Rotenberg, E. Sm–Nd systematics of chondrites. *Earth Planet. Sci. Lett.* **223**, 267–282 (2004).
44. Barrat, J. A. *et al.* Geochemistry of CI chondrites: major and trace elements, and Cu and Zn isotopes. *Geochim. Cosmochim. Acta* **83**, 79–92 (2012).



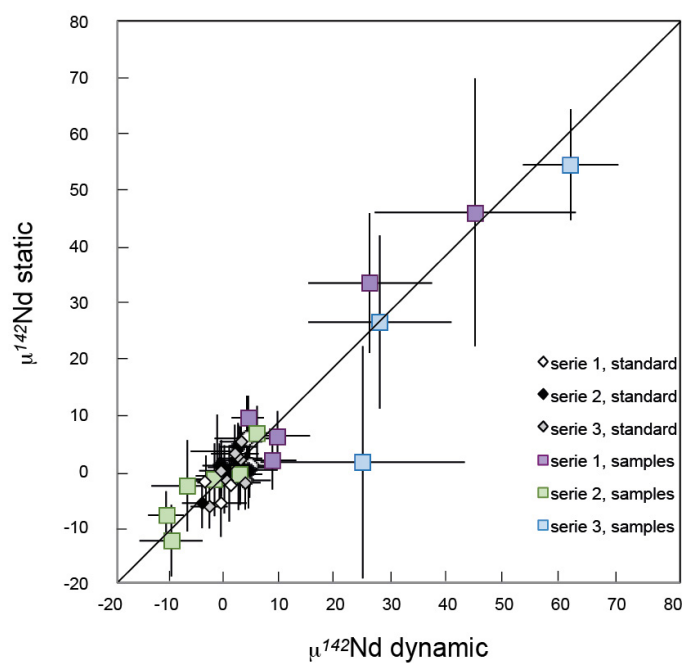
Extended Data Figure 1 | CI-normalized REE abundances in the four individual CAIs. CI average composition is from ref. 44. See Extended Data Table 1 for elemental concentrations.



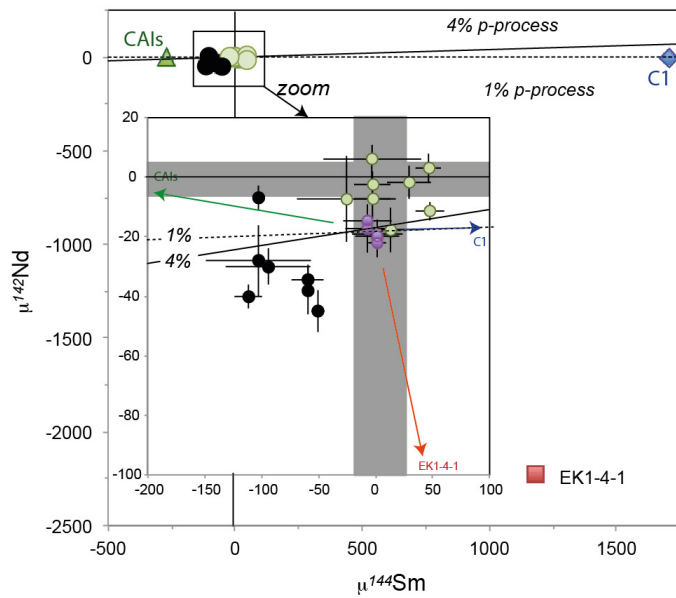
Extended Data Figure 2 | $^{146}\text{Nd}/^{147}\text{Sm}$ versus $\mu^{144}\text{Sm}$ from Allende 322 and 323, NWA 2364 and NWA 6991 bulk CAIs. The deviations are given in parts per million of $^{144}\text{Sm}/^{152}\text{Sm}$ ratios measured for samples compared to the average of measured Sm isotopic standards. Error bars represent internal errors (2 s.e.) for individual measurements. There is no correlation between the abundance of Nd within the Sm cuts with $\mu^{144}\text{Sm}$ compositions in individual CAIs. See Supplementary Tables 1–3 for isotopic data.



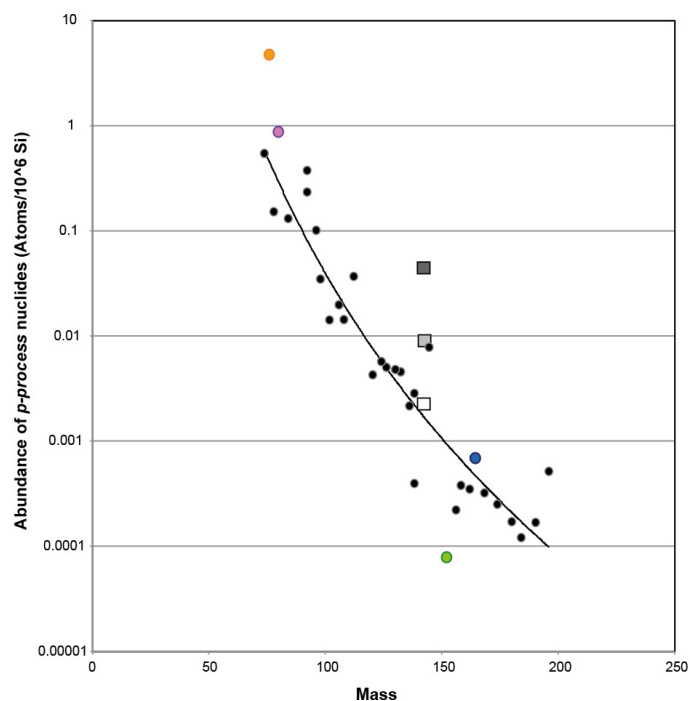
Extended Data Figure 3 | Sm isotopic ratios of CAIs and CV3 chondrites. Data for CV3 chondrites is from ref. 3. WR, whole rocks. Error bars indicate internal errors (2 s.e.) on individual measurements and black lines are best-fit lines.



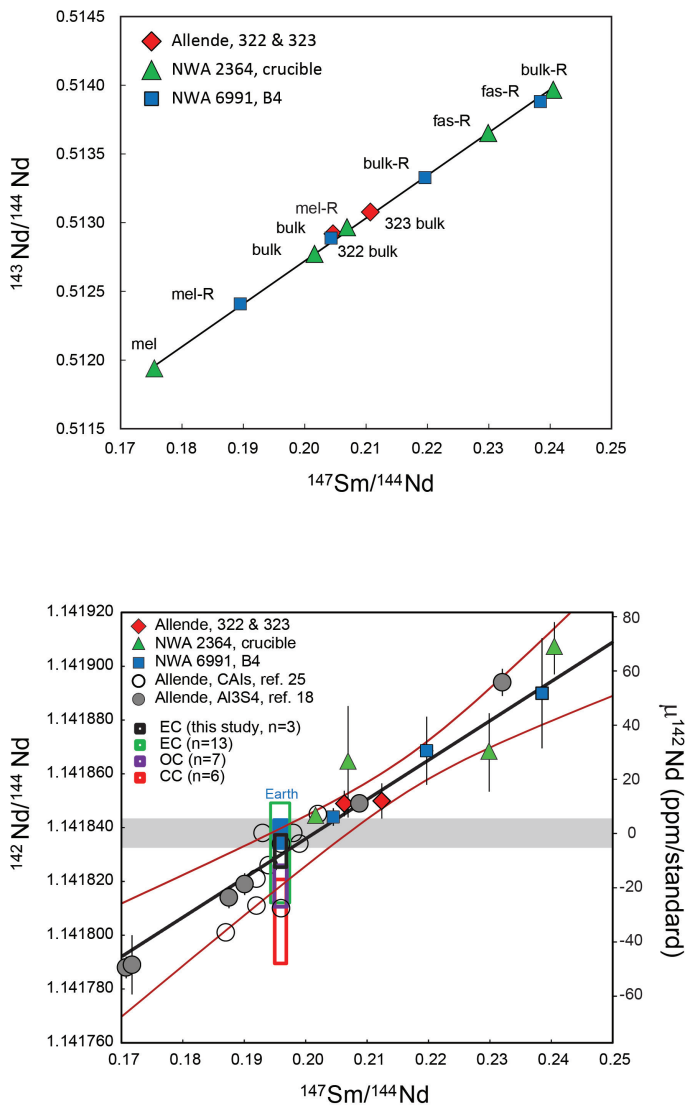
Extended Data Figure 4 | Deviation of $^{142}\text{Nd}/^{144}\text{Nd}$ ratios relative to the JNdi-1 standard measured in dynamic and static modes. Deviations are given in parts per million. For static mode, the Faraday cup for line 1 was centred at an atomic mass of 145. Error bars indicate internal errors (2 s.e.) on individual measurements and the black line indicates a slope of one.



Extended Data Figure 5 | $^{144}\text{Sm}/^{152}\text{Sm}$ versus $^{142}\text{Nd}/^{144}\text{Nd}$ expressed in μ notation. Chondrite data (black circles, carbonaceous chondrites; purple circles, ordinary chondrites; green circles, enstatite chondrites) are from this study and refs 3 and 4. CAIs, FUN inclusions (EK1-4-1 and C1; refs 41, 42) and chondrite $^{142}\text{Nd}/^{144}\text{Nd}$ ratios are corrected for radiogenic decay of ^{142}Nd over the age of the Solar System. Gray boxes show 2σ external reproducibility obtained on the standard. Solid and dotted lines correspond to p-process contributions for ^{142}Nd of 4% and 1%, respectively. The complementary part is formed by s-processes. Our data show that there is no correlation between $\mu^{144}\text{Sm}$ and $\mu^{142}\text{Nd}$, in contrast to previous suggestions⁴. Error bars indicate internal errors (2 s.e.) on individual measurements when larger than symbols and available.



Extended Data Figure 6 | Abundance of p-process nuclides versus mass. Black dots show isotopes formed by p-processes only; coloured dots show isotopes formed partly by p-process and partly by s-processes (orange, ⁷⁶Se; purple, ⁸⁰Kr; green, ¹⁵²Gd; blue, ¹⁶⁴Er). Model abundances of ¹⁴²Nd are represented by the squares coloured in dark grey for a 20% p-process contribution, light grey for 4% and white for 1%. The black line is the best-fit line.



Extended Data Figure 7 | Sm-Nd internal isochrons of CAIs from

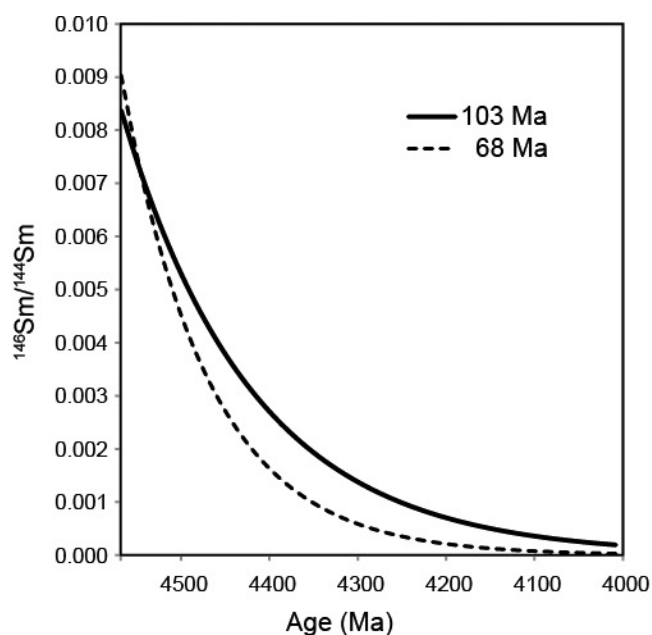
Allende, NWA 2364 and NWA 6991. **a**, ^{147}Sm - ^{143}Nd internal isochron.

By combining all the CAI fractions together and fitting the data to a straight line, we determine the ^{147}Sm - ^{143}Nd age to be $4,526 \pm 150$ Myr (MSWD = 1.2, initial $^{143}\text{Nd}/^{144}\text{Nd} = 0.50673 \pm 0.00021$). R, residue after leaching; fas, fassaite; mel, melilite. **b**, The black line represents the

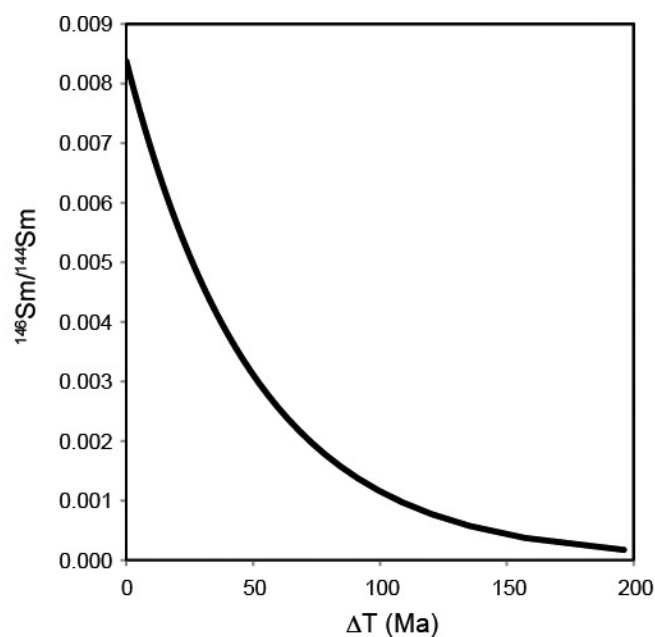
^{146}Sm - ^{142}Nd internal isochron of CAIs from Allende 322 and 323, NWA 2364 and NWA 6991; red lines indicate the 95% confidence interval.

^{146}Sm - ^{142}Nd systematics of Allende bulk and mineral separates (Al3S4)¹⁸ and Allende bulk CAIs²⁵ are shown for comparison (error bars are not shown because individual analytical errors were not provided).

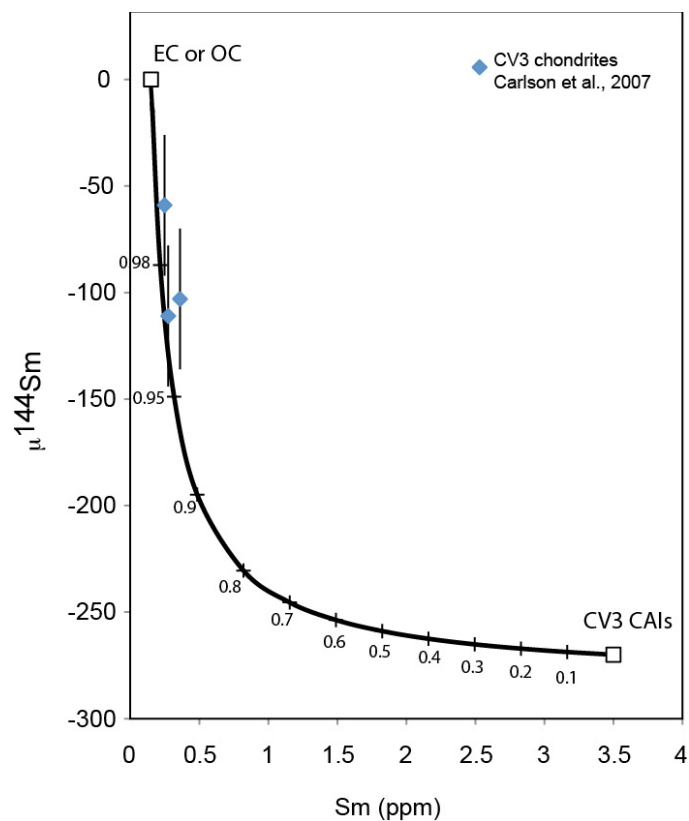
The blue rectangle represents the composition of modern Earth's mantle, as represented by our long-term measurements of the JNdi-1 standard ($^{142}\text{Nd}/^{144}\text{Nd} = 1.141838 \pm 0.000006$; by definition, $\mu^{142}\text{Nd} = 0$) with $^{147}\text{Sm}/^{144}\text{Nd} = 0.1960$, which is within the error of the regression for CAIs from this study. The other rectangles represent the averages for $\mu^{142}\text{Nd}$ with 2 s.d. given below with values of -7 ± 6 p.p.m. for enstatite chondrites (EC; black, value from this study, $n = 3$), -6 ± 18 p.p.m. for enstatite chondrites (EC; green, $n = 14$), -18 ± 6 p.p.m. for ordinary chondrites (OC; purple, $n = 5$) and -34 ± 18 p.p.m. for carbonaceous chondrites (CC; red, $n = 8$), all normalized at $^{147}\text{Sm}/^{144}\text{Nd} = 0.1960$ (the widths of the rectangles for $^{147}\text{Sm}/^{144}\text{Nd}$ are exaggerated for clarity).



Extended Data Figure 8 | Evolution of the $^{146}\text{Sm}/^{144}\text{Sm}$ ratio. Left, $^{146}\text{Sm}/^{144}\text{Sm}$ ratio as a function of age using the two proposed decay constants for the short-lived ^{146}Sm radionuclide. The two curves intersect at $^{146}\text{Sm}/^{144}\text{Sm} = 0.0073$ and 4,546 Ma—the values defined from eucrite



internal isochron of Binda in ref. 8. For objects formed after 4,546 Ma ago, ages are younger using the 103-Ma half-life value. Right, time difference ΔT obtained using the two decay constants proposed for ^{146}Sm . Ma, millions of years.



Extended Data Figure 9 | Mixing model (black line) between CV3 CAIs (^{144}Sm abundance measured in this study) and matrix (enstatite chondrite (EC) and ordinary chondrite (OC) whole-rock meteorites without a ^{144}Sm anomaly) (open squares). ^{144}Sm anomalies measured in CV3 chondrites (blue symbols) correspond to 1%–3% CAI volume abundances in the mixing model^{3,20}. The numbers on the line indicate the proportion of matrix-component end-member relative to CAI end-member. Error bars (2 s.e.) are for individual measurements of whole-rock CV3 chondrite meteorites.

Extended Data Table 1 | Concentrations of REEs in individual bulk and unleached CAIs

Method	Samples	La	Ce	Pr	Nd	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
ID	Allende, 322	-	-	-	9.00	3.02	-	-	-	-	-	-	-	-	0.007
ID	Allende, 323	-	-	-	9.47	3.27	-	-	-	-	-	-	-	-	0.070
ID	NWA 2364, crucible	-	-	-	7.56	2.51	-	-	-	-	-	-	-	-	0.490
ICPMS	NWA 6991, B4	4.12	10.9	1.62	8.73	3.03	1.32	3.92	0.691	4.93	1.01	3.38	0.448	3.25	0.475

Concentrations are given in parts per million and were measured by quadrupole ICPMS or isotopic dilution (ID) methods by MC-ICPMS.

Discovery of species-wide tool use in the Hawaiian crow

Christian Rutz¹, Barbara C. Klump¹, Lisa Komarczyk², Rosanna Leighton², Joshua Kramer³, Saskia Wischniewski¹, Shoko Sugawara¹, Michael B. Morrissey¹, Richard James⁴, James J. H. St Clair¹, Richard A. Switzer⁵ & Bryce M. Masuda²

Only a handful of bird species are known to use foraging tools in the wild¹. Amongst them, the New Caledonian crow (*Corvus moneduloides*) stands out with its sophisticated tool-making skills^{2,3}. Despite considerable speculation, the evolutionary origins of this species' remarkable tool behaviour remain largely unknown, not least because no naturally tool-using congeners have yet been identified that would enable informative comparisons⁴. Here we show that another tropical corvid, the 'Alalā (*C. hawaiiensis*; Hawaiian crow), is a highly dexterous tool user. Although the 'Alalā became extinct in the wild in the early 2000s, and currently survives only in captivity⁵, at least two lines of evidence suggest that tool use is part of the species' natural behavioural repertoire: juveniles develop functional tool use without training, or social input from adults; and proficient tool use is a species-wide capacity. 'Alalā and New Caledonian crows evolved in similar environments on remote tropical islands, yet are only distantly related⁶, suggesting that their technical abilities arose convergently. This supports the idea that avian foraging tool use is facilitated by ecological conditions typical of islands, such as reduced competition for embedded prey and low predation risk^{4,7}. Our discovery creates exciting opportunities for comparative research on multiple tool-using and non-tool-using corvid species. Such work will in turn pave the way for replicated cross-taxonomic comparisons with the primate lineage, enabling valuable insights into the evolutionary origins of tool-using behaviour.

The foraging behaviour of many corvid species remains poorly studied⁸, so it is possible that there are undiscovered tool users in this genus⁴. We identified the 'Alalā as a promising candidate for further investigation (see p. 161 in ref. 4), on the basis of its morphological^{9,10} and ecological⁴ similarity to the tool-using New Caledonian crow (Fig. 1, c, d and Extended Data Fig. 1a). Following a precipitous decline in the late twentieth century⁵, the world's entire 'Alalā population currently resides in two captive facilities where birds are being bred for future release¹¹ (Figs 1f and 2b). After studying anecdotal reports^{12,13}, the instigating authors learned from facility staff that tool use had indeed been repeatedly observed over the years (Supplementary Video 4; see Methods), leading to the collaborative project reported here.

We tested 104 of the 109 surviving 'Alalā (five birds were excluded *a priori* for health reasons), and found that 78% of birds spontaneously used tools to probe for out-of-reach food (Fig. 2f). While tool-use competence (that is, whether or not a bird used tools) was very similar between males and females (Fig. 2c), competence varied strongly across age classes (Fig. 2d): 93% of all sexually mature subjects (third year of life or older⁵) were confirmed as tool users, compared to 47% of younger birds. In the majority of cases, birds used tools in their very first trial, usually within minutes of gaining access to the experimental apparatus, a wooden log with six extraction tasks (Fig. 2a and Extended Data Fig. 2a). Most subjects handled stick tools in a highly dexterous

manner (Supplementary Videos 1, 2) and extracted bait from several tasks (median 4, range 0–6; $n = 64$ individually tested tool users). All but one successful extractions from vertical and horizontal crevices and drilled horizontal holes were completed in less than 60 s of probing time, with vertical holes proving slightly more challenging (Fig. 2g). During experimental trials, birds routinely selected tools of appropriate dimensions, replaced unsuitable tools, and transported non-supplied sticks to the log. Tool modification occurred frequently (shortening: 67% of $n = 64$ individually tested tool users; other modifications: 8%), and we even observed tools being manufactured from plant materials (14%) (Supplementary Video 2). 'Alalā have relatively straight bills and highly mobile eyes (Extended Data Fig. 1 and Supplementary Video 5)—features that are thought to facilitate dexterous handling of bill-held tools in New Caledonian crows^{9,14} (for craniofacial morphology of other extant crows and two extinct Hawaiian species, see Fig. 1b, f).

Our discovery of a species-wide capacity for tool use raises the possibility that 'Alalā possess genetic predispositions similar to those reported for New Caledonian crows^{15,16}. To examine this hypothesis, we reared seven naive juvenile 'Alalā in two social groups under controlled conditions, without opportunities to observe tool-proficient adults. All birds eventually used sticks and other objects in an attempt to reach hidden food during probe trials (Fig. 3b, Extended Data Fig. 2b and Extended Data Table 1), and four were successful (Fig. 3c and Supplementary Video 3; a fifth subject later used tools successfully on the log task). Towards the end of the 5-month observation period, we documented an increase in the handling of stick-type and similar objects (Fig. 3a), possibly in response to increased exposure to tool-use opportunities (Fig. 3c), but 'Alalā did not perform the stereotyped probing or rubbing behaviours that are precursors of functional tool use in New Caledonian crows¹⁶. 'Alalā also appeared to spend less time manipulating stick-type and similar objects 3–5 weeks post-fledging than New Caledonian crows, with some estimates even lower than for non-tool-using ravens (*C. corax*)¹⁷ (Fig. 3d), although these comparisons should be treated cautiously owing to differences in study protocols.

While our rearing experiment demonstrated conclusively that naive 'Alalā can independently develop functional tool use, environmental conditions are likely to affect behavioural development. At the population level, we detected only minor differences between birds that had been raised (and tested) at the two facilities (Fig. 2c), despite some variation in enrichment regimes. In groups of young 'Alalā, we often observed birds interfering with each other's attempts to use tools, for example by stealing sticks (Supplementary Video 3). We examined possible social-interference effects in a separate experiment, in which birds were tested both in their usual housing group (of 6–7 subjects) and individually. Tool-use behaviour was generally rare amongst 'Alalā in their second year of life, irrespective of experimental condition, but it was clearly suppressed by the presence of group mates in subjects that were a year older (Fig. 2e).

¹Centre for Biological Diversity, School of Biology, University of St Andrews, Sir Harold Mitchell Building, St Andrews KY16 9TH, UK. ²Institute for Conservation Research, San Diego Zoo Global, PO Box 39, Volcano, Hawai'i 96785, USA. ³Institute for Conservation Research, San Diego Zoo Global, 2375 Olinda Road, Makawao, Hawai'i 96768, USA. ⁴Department of Physics and Centre for Networks and Collective Behaviour, University of Bath, Bath BA2 7AY, UK. ⁵Institute for Conservation Research, San Diego Zoo Global, 15600 San Pasqual Valley Road, Escondido, California 92027, USA.

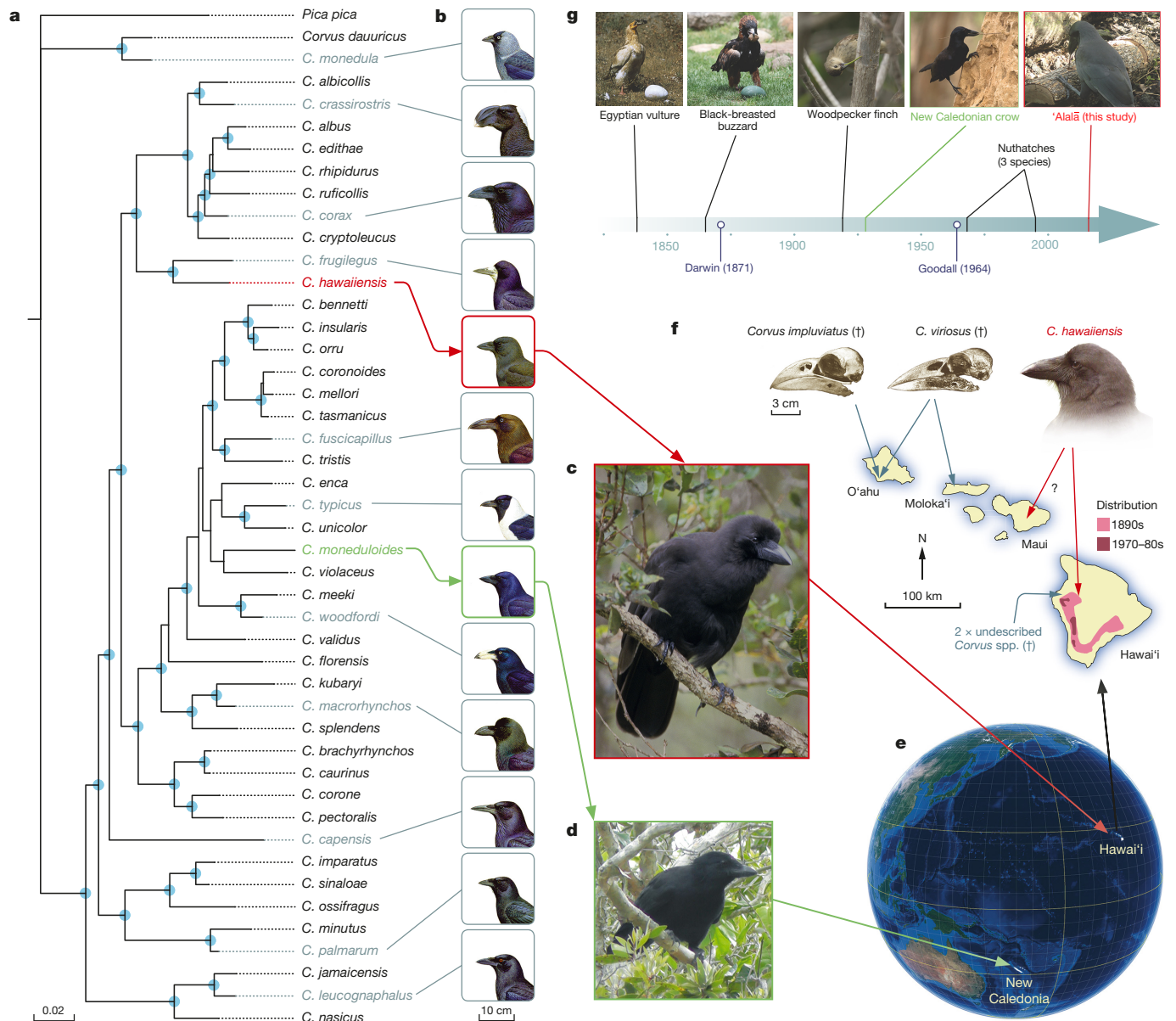


Figure 1 | Phylogenetic and biogeographical context of tool behaviour in crows. **a**, Phylogeny for the genus *Corvus* (blue circles, posterior probabilities ≥ 0.90). Scale bar: estimated substitutions per site. **b**, Variation in craniofacial morphology (adapted from ref. 8, Lynx Edicions). **c**, One of the last wild 'Alalā (27 February 1998, Kealakekua, Hawai'i; photo: Jack Jeffrey Photography). **d**, New Caledonian crow (photo: M. Griffioen). **e**, Location of Hawai'i and New Caledonia (globe: Google Earth,

NASA, US Geological Survey). **f**, Hawaiian corvids (skulls adapted from ref. 26, American Ornithologists' Union; photo: C.R.), and historical 'Alalā distribution (from ref. 11, USFWS). † denotes extinct species. **g**, Discovery timeline for well-known habitual avian tool users (photos: A. Gandolfi/naturepl.com; D. Pintimalli; D. Brinkhuizen; J. Troscianko), with landmark chimpanzee reports by Darwin²⁷ and Goodall²⁸ for reference.

Using detailed housing data and computer simulations, we next examined the social connectivity of our study population, by tracing potential transmission pathways (Fig. 3e, right) in time-ordered contact networks (1996–2013; Fig. 3e, left). On the basis of highly conservative assumptions (instantaneous, deterministic information transfer), we estimated that between one (unrestricted transmission) and eight (more realistic, age-biased transmission¹⁸) independent information sources would be required to reach all confirmed tool users by 2013 ($n = 74$ birds, excluding the 7 isolated subjects of our rearing experiment). This indicates that, despite considerable social mixing, it is unlikely that a single 'innovation' event can explain the observed species-wide distribution of tool competence. 'Alalā clearly possess a propensity to 'discover' tool-assisted foraging solutions independently, which probably results from genetically canalised, persistent object-exploration behaviour; further experiments are

required to quantify the relative contributions of individual and social learning¹⁹.

It is well known that naturally non-tool-using animal species sometimes use tools in captivity, especially when the behaviour is shaped or otherwise encouraged¹. The case of the 'Alalā is unusual in several regards: almost all adult birds expressed tool behaviour (Fig. 2c); tool users swiftly solved even demanding extraction tasks (Fig. 2g); and naive subjects independently acquired tool-using skills (Fig. 3c). Comparison with naturally non-tool-using corvids reveals another difference. Most 'Alalā and New Caledonian crows exhibit a striking degree of dexterity during stick handling, while captive rooks (*C. frugilegus*) appear to have less control over their tools²⁰ (Supplementary Video 6). We observed rook-like tool handling in the seven juveniles of our rearing experiment, but this was unusual amongst older 'Alalā, suggesting that tool control improves with practice; we note, however,

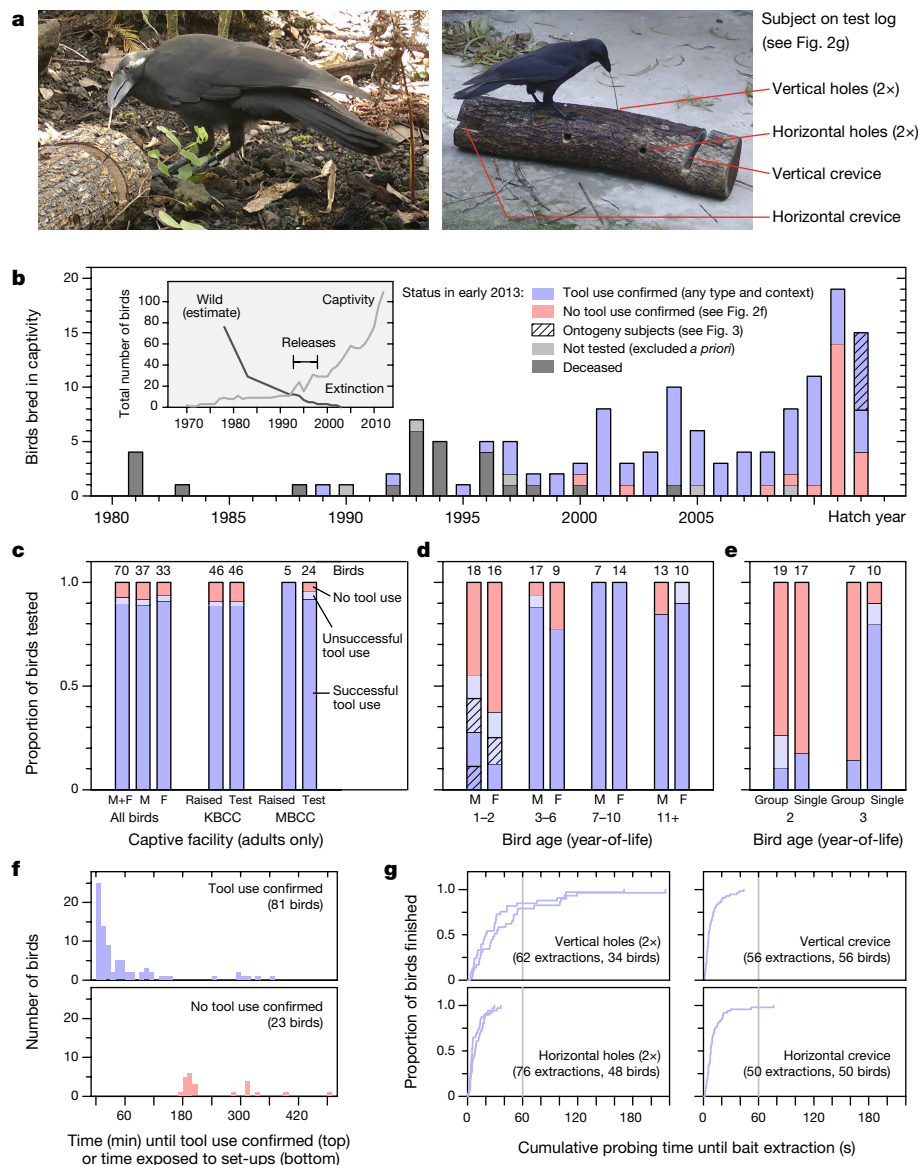


Figure 2 | Species-wide tool-use behaviour in 'Alalā. **a**, Captive birds using stick tools to extract bait from experimental logs. **b**, Development of the world's 'Alalā population and results of species-wide tool-use assay (birds shown survived at least until post-fledging age; inset data from ref. 29, Elsevier). **c–e**, Tool-use competence across: **c**, sexes (M, male; F, female) and facilities (KBCC, Keauhou Bird Conservation Center; MBCC, Maui Bird Conservation Center); **d**, age classes; and **e**, different test conditions (tested individually or in a group). **f**, **g**, Bird performance: **f**, outcome of trials and **g**, extraction speed for different tasks. Panels **b–g** refer to the standardized tool-use assay (Extended Data Fig. 2a); **g** only includes successful extractions from the first individual trial where birds used tools.

that even highly proficient adults would have had relatively limited tool-use experience during their lifetimes.

'Alalā once lived in dry- and wet-forest habitats on Hawai'i Island (Fig. 1f), where they foraged for a variety of fruit, invertebrates and other items^{5,21}. Wild birds have been observed to engage in woodpecker-like extractive foraging, flaking bark and chiselling wood with their powerful bills^{5,21,22}, just as New Caledonian crows are known to do^{4,23}. Apart from one suggestive observation, however, of a bird transporting a twig in its bill (P. Crosland, cited in ref. 22)—at a time of year (late June) when nest construction was unlikely⁵—we have found no reports of tool-related behaviour in the wild. Tool use may have been relatively infrequent, confined to particular habitats, or difficult to observe (Extended Data Table 2). Alternatively, the last wild 'Alalā may have no longer used tools, for example, if island-wide habitat degradation²⁴ had forced them to switch to alternative foraging modes—a scenario with important implications for forthcoming reintroduction attempts¹¹.

Anecdotal observations of avian tool use are relatively common, but very few species routinely use foraging tools in the wild¹

(for well-known examples, see Fig. 1g). Unfortunately, because the 'Alalā is extinct in the wild, and tools made from plant materials are perishable, we may never know whether these birds once used tools under natural conditions. Current evidence strongly favours this scenario, but otherwise, our study would have uncovered a truly remarkable capacity for highly dexterous tool behaviour in a naturally non-tool-using corvid. Future studies should chart the (development of) object-related behaviour in other species under similar conditions in captivity, with an initial focus on the rook, which is the 'Alalā's sister species⁶ (Fig. 1a), and a rapid learner of tool skills, when trained appropriately^{20,25}.

'Alalā and New Caledonian crows are only distantly related¹⁶ (Fig. 1a), suggesting that tool-related adaptations evolved convergently. In fact, interspecific differences in the ontogenetic development of functional tool use support the hypothesis of convergence rather than homology. As for possible ecological drivers, both species^{4,26}—as well as the stick-tool-using Galápagos woodpecker finch⁷—evolved on remote tropical islands (Fig. 1e) where competition for embedded prey is likely to be reduced and predation risk low. These conditions, which have

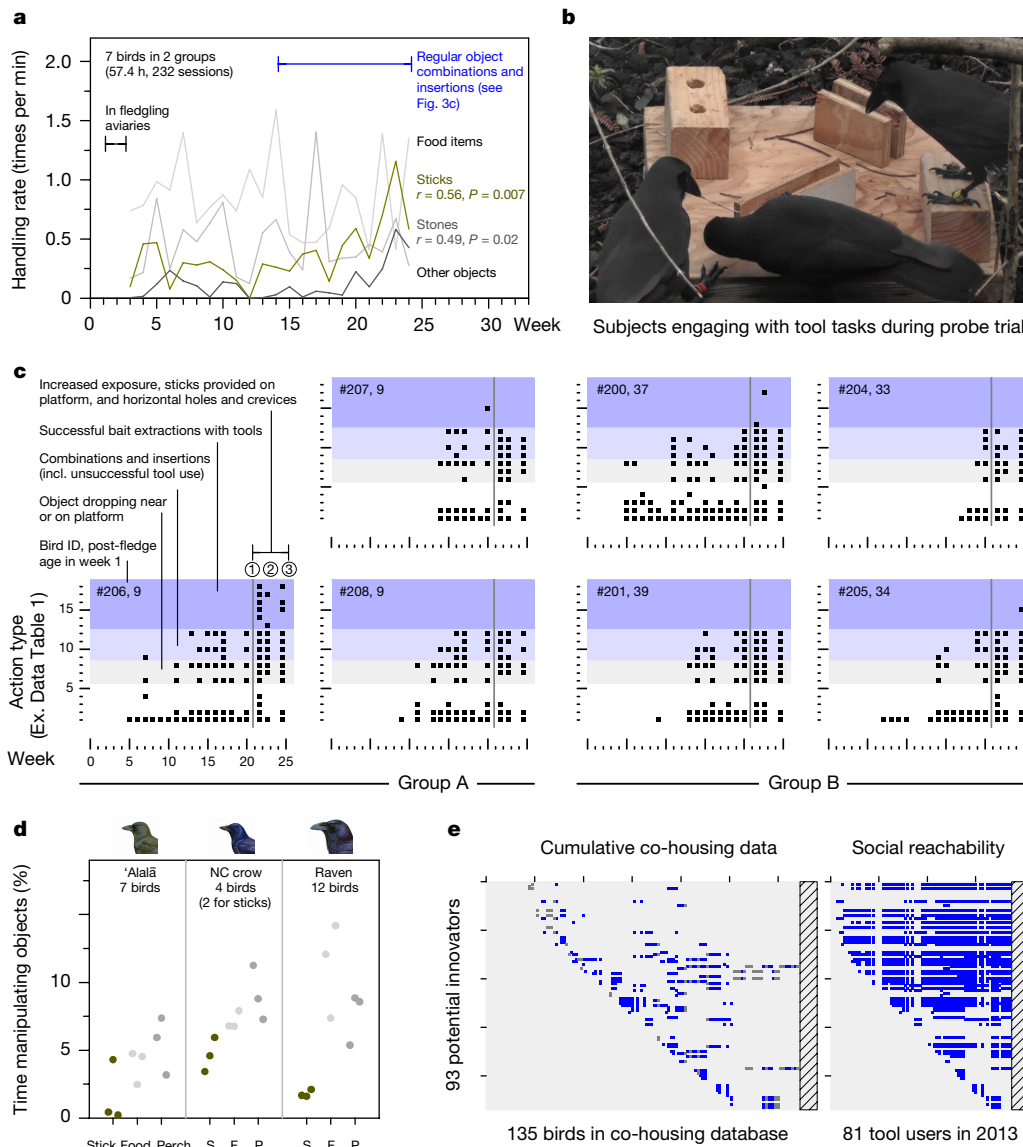


Figure 3 | Development of tool-use behaviour in naive, juvenile 'Alalā. **a**, Object-handling rates (bill only) estimated from focal-bird observations (week 1 commenced 3 September 2012; 'sticks' are all stick-type objects, fern sections and branched pieces of plant; correlation coefficients). **b**, Group A on experimental platform. **c**, Behavioural development as documented through weekly probe trials, from week 3 onward (action types are defined in Extended Data Table 1; for ①–③, see Methods).

previously been predicted to facilitate tool behaviour^{4,7}, may vary across island environments, but are presumably less common on adjacent mainland habitats, providing a possible explanation for the striking rarity of avian tool use¹.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 February; accepted 12 July 2016.

- Shumaker, R. W., Walkup, K. R. & Beck, B. B. *Animal Tool Behavior: the Use and Manufacture of Tools by Animals* (Johns Hopkins Univ. Press, 2011).
- Hunt, G. R. Manufacture and use of hook-tools by New Caledonian crows. *Nature* **379**, 249–251 (1996).
- Klump, B. C., Sugasawa, S., St Clair, J. J. H. & Rutz, C. Hook tool manufacture in New Caledonian crows: behavioural variation and the influence of raw materials. *BMC Biol.* **13**, 97 (2015).
- Rutz, C. & St Clair, J. J. H. The evolutionary origins and ecological context of tool use in New Caledonian crows. *Behav. Processes* **89**, 153–165 (2012).

d, Comparison of object-manipulation times (bill, and foot-grasped¹⁷) 3–5 weeks post-fledging (weekly means; note that $n < 5$ birds for some values) between 'Alalā, New Caledonian (NC) crows¹⁷ and non-tool-using ravens¹⁷. **e**, Potential for social diffusion across the 'Alalā contact network (1996–2013, cumulative results; hatching blocks out the 7 isolated subjects of the rearing experiment): co-housing matrix (left) and reachability matrix (right).

- Banko, P. C., Ball, D. L. & Banko, W. E. in *The Birds of North America*, vol. 17 (eds Poole, A. & Gill, F.) (The Birds of North America, Inc., 2002).
- Haring, E., Däubel, B., Pinsker, W., Kryukov, A. & Gamauf, A. Genetic divergences and intraspecific variation in corvids of the genus *Corvus* (Aves: Passeriformes: Corvidae) – a first survey based on museum specimens. *J. Zool. Syst. Evol. Res.* **50**, 230–246 (2012).
- Tebich, S. & Teschke, I. in *Tool Use in Animals: Cognition and Ecology* (eds Sanz, C. M., Call, J. & Boesch, C.) (Cambridge Univ. Press, 2013).
- del Hoyo, J., Elliott, A. & Christie, D. A. (eds) *Handbook of the Birds of the World. Bush-shrikes to Old World Sparrows*, vol. 14 (Lynx Edicions, 2009).
- Troscianko, J., von Bayern, A. M. P., Chappell, J., Rutz, C. & Martin, G. R. Extreme binocular vision and a straight bill facilitate tool use in New Caledonian crows. *Nat. Commun.* **3**, 1110 (2012).
- Goodwin, D. *Crows of the World* (British Museum, 1976).
- U.S. Fish and Wildlife Service (USFWS). *Revised Recovery Plan for the 'Alalā (Corvus hawaiiensis)* (USFWS, 2009).
- Faike, E. Wild voices in captivity: the fate of the 'Alalā. *Birding* **38**, 64–67 (2006).
- Smith, M. 'Alalā: weighing in. <http://blog.sandiegozooglobal.org/2011/10/21/alala-weighing-in/> (2011).
- Matsui, H. et al. Adaptive bill morphology for enhanced tool manipulation in New Caledonian crows. *Sci. Rep.* **6**, 22776 (2016).

15. Kenward, B., Weir, A. A. S., Rutz, C. & Kacelnik, A. Tool manufacture by naive juvenile crows. *Nature* **433**, 121 (2005).
16. Kenward, B., Rutz, C., Weir, A. A. S. & Kacelnik, A. Development of tool use in New Caledonian crows: inherited action patterns and social influences. *Anim. Behav.* **72**, 1329–1343 (2006).
17. Kenward, B. *et al.* On the evolutionary and ontogenetic origins of tool-oriented behaviour in New Caledonian crows (*Corvus moneduloides*). *Biol. J. Linn. Soc.* **102**, 870–877 (2011).
18. Mesoudi, A., Chang, L., Dall, S. R. X. & Thornton, A. The evolution of individual and cultural variation in social learning. *Trends Ecol. Evol.* **31**, 215–225 (2016).
19. Tennie, C., Call, J. & Tomasello, M. Ratcheting up the ratchet: on the evolution of cumulative culture. *Phil. Trans. R. Soc. B* **364**, 2405–2415 (2009).
20. Bird, C. D. & Emery, N. J. Insightful problem solving and creative tool modification by captive nontool-using rooks. *Proc. Natl Acad. Sci. USA* **106**, 10370–10375 (2009).
21. Sakai, H. F., Ralph, C. J. & Jenkins, C. D. Foraging ecology of the Hawaiian crow, an endangered generalist. *Condor* **88**, 211–219 (1986).
22. Walters, M. J. *Seeking the Sacred Raven* (Island Press, 2006).
23. Troscianko, J. & Rutz, C. Activity profiles and hook-tool use of New Caledonian crows recorded by bird-borne video cameras. *Biol. Lett.* **11**, 20150777 (2015).
24. Pratt, T. K., Atkinson, C. T., Banko, P. C., Jacobi, J. D. & Woodworth, B. L. (eds) *Conservation Biology of Hawaiian Forest Birds* (Yale Univ. Press, 2009).
25. Tebbich, S., Seed, A. M., Emery, N. J. & Clayton, N. S. Non-tool-using rooks, *Corvus frugilegus*, solve the trap-tube problem. *Anim. Cogn.* **10**, 225–231 (2007).
26. James, H. F. & Olson, S. L. Descriptions of thirty-two new species of birds from the Hawaiian Islands: Part II. Passeriformes. *Ornithol. Monogr.* **46**, 1–88 (1991).
27. Darwin, C. R. *The Descent of Man, and Selection in Relation to Sex* (John Murray, 1871).
28. Goodall, J. Tool-using and aimed throwing in a community of free-living chimpanzees. *Nature* **201**, 1264–1266 (1964).
29. Hoeck, P. E. A., Wolak, M. E., Switzer, R. A., Kuehler, C. M. & Lieberman, A. A. Effects of inbreeding and parental incubation on captive breeding success in Hawaiian crows. *Biol. Conserv.* **184**, 357–364 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank: the many people – far too many to name individually – who have prevented the ‘Alalā’s extinction, and who are working tirelessly towards its successful reintroduction into the wild; R. Fleischer for facilitating initial contacts; San Diego Zoo Global’s staff for assistance with experiments; C. Higgott for help with video scoring; D. Parker for constructing the consensus phylogeny; several photographers for providing images for Fig. 1; S. Thompson for help with graphic design; and G. Ruxton for statistical advice. Research was conducted with permission from San Diego Zoo Global’s IACUC animal welfare committee (Project ID#12-017), and with funding from the Biotechnology and Biological Sciences Research Council, UK (BBSRC; grant BB/G023913/2 to C.R., and studentship to B.C.K.), the University of St Andrews (C.R.), JASSO (S.S.), and the Royal Society of London (M.B.M.). Funding for the captive ‘Alalā propagation programme was provided by the U.S. Fish and Wildlife Service, Hawai’i Division of Forestry and Wildlife, Moore Family Foundation, Marisla Foundation, several anonymous donors, and San Diego Zoo Global.

Author Contributions C.R. conceived of, initiated and led the project, and secured funding; R.A.S. and B.M.M. led the captive ‘Alalā propagation programme, with support from R.L., J.K. and L.K.; C.R., J.J.H.S.C. and B.C.K. designed behavioural experiments; B.C.K., J.J.H.S.C. and C.R. conducted the species-wide tool-use assay, and L.K. performed the ontogeny experiment, with help from R.L., C.R., J.J.H.S.C. and B.C.K.; B.C.K. scored videos and extracted behavioural data, except for ontogeny trials, which were scored by S.W. and processed by S.S.; L.K., J.K., B.M.M. and C.R. collated and checked data on the life histories of captive birds; M.B.M. and B.C.K. assessed bird-performance data; R.J. and C.R. conducted social-diffusion simulations; C.R. and B.C.K. coordinated manuscript preparation, analysed data and prepared figures and videos; and C.R. wrote the manuscript, which was edited by B.C.K., J.J.H.S.C., R.J. and B.M.M., and approved by all co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.R. (christian.rutz@st-andrews.ac.uk).

Reviewer Information *Nature* thanks T. Bugnyar, B. Kenward, S. Tebbich and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

The use of statistical methods to predetermine sample sizes was not necessary: in the main experiment, all healthy individuals of the world's 'Alalā population were tested, and all other experiments (as detailed below) likewise attempted to maximise sample sizes. Randomization procedures were used to establish the order in which subjects were observed in some experiments (as detailed below), and the order in which all video files were analysed; all videos for the assessment of inter-observer agreement were randomly selected. For some video analyses (as detailed below), scorers were hypothesis naïve.

Study population. 'Alalā were studied in two captive breeding facilities operated by San Diego Zoo Global. With the species considered extinct in the wild^{11,30}, the world's population consisted of 109 individuals (58 males; 51 females) in early 2013, with: 64 birds housed at the Keauhou Bird Conservation Center (KBCC), Hawai'i Island; 44 birds at the Maui Bird Conservation Center (MBCC), Maui; and a single individual off-exhibit at San Diego Zoo Safari Park, California. The captive stock originated from a few founder individuals that had been collected from the wild since the 1970s, as described in detail elsewhere^{11,29–32}. All birds available for testing in our study (referred to throughout by their studbook numbers) were of known ancestry, sex (determined through genetic analysis of blood samples³³) and age, and had been reared in captivity (Fig. 2b). Male #67 had hatched from one of the last eggs laid by a wild pair, and three other subjects (#77, #78, #86) had temporarily lived in the wild (they had been released in the late 1990s, but were later returned to captivity³⁰).

Adult birds were kept as breeding pairs, or sometimes as singletons, and immature birds were housed in groups of up to eight individuals, to facilitate their socialization³⁴. All aviaries at the two main facilities are multi-chambered, spacious outdoor enclosures (varying in size from $\sim 3.0 \times 6.0 \times 3.7$ m to $7.3 \times 17.0 \times 5.5$ m), which are open to the elements, but have a roofed section for shelter. At the KBCC (purpose-built in 1996), the ground is covered in lava stones, with patches of live vegetation, while at the MBCC (repurposed building in use since 1986, with later extensions), some aviaries have concrete flooring. Standard fittings include a variety of branches and ropes for perching, a nesting platform, and a large water bath. All birds have access to cut vegetation ('browse') and sticks year-round, and pairs receive supplies of assorted nesting material during the breeding season.

Enrichment protocols have changed over the years and varied slightly between facilities. Initially, all enrichment given to 'Alalā was made of natural materials (for example, fresh browse, and logs of deadwood), but this was supplemented with artificial items (for example, food hidden inside dog toys, or wrapped in newspaper) from 2008 at the KBCC (and at the latest from 1999 onwards at the MBCC); a human-imprinted male (#35) was given artificial items as early as 2000. Food items were hidden in holes and crevices in wooden logs, or tossed into water baths, intermittently since at least 1997, and about once or twice a week since 2004, at the KBCC (since 1999 at the MBCC), and baited PVC tubes were presented from late 2012 onwards (since 2007 at MBCC). While this enrichment provided opportunities for tool use, in the vast majority of cases bait could also be obtained by bill alone, in contrast to the extraction tasks of our formal behavioural assay (see below). Importantly, to the best of our knowledge, the use of tools to extract hidden food was never demonstrated to birds at either facility.

Behavioural assay. We conducted a species-wide assay of tool-use competence, using a standardized food-extraction task set (see below). Following pilot experiments with two subjects (female #94, and her son #134) in August 2012 and January 2013, we tested all healthy birds in both facilities between 23 January and 27 February 2013. With five birds excluded from experiments *a priori* for medical reasons, and one male tested later in the year (#67; tool use confirmed on 31 August 2013), our final sample comprised 104 subjects, which was over 95% of the world's 'Alalā population at the time (see Fig. 2b). As we effectively tested an entire species, it was not necessary to use inferential statistics to support findings.

The experimental set-up consisted of (Extended Data Fig. 2a): a Koa *Acacia koa* log containing four drilled holes and two crevices, each baited with a quarter of a neonate mouse (or other preferred food in early trials at KBCC); 12 sticks of varying lengths as potential tools scattered in front of the log; and assorted native plant materials (KBCC), or two dead branched stems (MBCC; native materials not readily available), from which tools could be manufactured, wedged firmly into a wooden board to stand upright (for further details, see Extended Data Fig. 2a). The four different types of extraction task were designed to resemble foraging problems New Caledonian crows regularly solve with tools in the wild^{24,23}. At both facilities, we used the same two near-identical logs to run trials in parallel. Encouraged by earlier anecdotal observations during routine enrichment sessions (Supplementary Video 4), we usually also placed a piece of mouse head in the aviary's water bath, to see whether the subject(s) would fish it out with a stick; this complementary task proved useful, as it often attracted birds' attention, and confirmed tool-use behaviour in one female (#95) that failed to engage with the main log set-up.

Trials were scheduled to last for ~ 1.0 – 1.5 h, but were terminated earlier on a few occasions at the start of the study, while the test protocol was being established ($n = 6$ trials), or when all bait had been extracted ($n = 24$), cameras failed ($n = 2$) or due to experimenter error ($n = 1$). Food bowls were usually removed shortly before trials commenced, but birds sometimes found food scraps in their aviaries, and always had *ad libitum* access to water. An experimenter placed the fully-baited experimental log and the board with plant materials on the ground, before scattering the sticks underneath a large cotton sheet, out of view of the subject(s). Before the experimenter removed the sheet and left the aviary, several small food items were conspicuously placed on top of the log, to encourage approach and exploration of the set-up, and the water bath was baited (see above). At the KBCC, birds could be filmed with experimenter-operated video cameras through tinted or one-way-mirror observation windows, while at the MBCC, all trials had to be filmed with static video cameras hidden inside a rainproof box, placed ~ 1.5 – 3.0 m away from the experimental set-up. Subjects were temporarily isolated for individual testing ($n = 83$ birds), but we also ran some trials with pairs early on in the study ($n = 3$ birds) and some with larger groups where isolation was impossible owing to aviary layout ($n = 18$ birds). For logistical and ethical reasons, birds remained in visual contact with other 'Alalā in adjacent chambers even when tested individually. Subjects that did not show tool-related behaviours in their first trial were re-tested for varying amounts of time (Fig. 2f). Immature 'Alalā are usually housed in groups (see above); to examine experimentally how social context affects the expression of tool behaviour, we tested a sample of birds in their second and third year of life, both in their usual housing group and individually (Fig. 2e).

Video footage from experimental trials was scored in randomized order by the same observer (B.C.K.) using Solomon Coder software³⁵, and a subsample of 10 trials was re-scored by a second observer (S.S.) to estimate inter-observer agreement (Cohen's κ for 'extraction type' [tool/bill/not-extracted] = 0.97, $n = 70$ cases; correlation coefficient r for 'time spent probing with a tool' = 0.99, $P < 0.0001$, $n = 38$ probing bouts); all analyses are based on the original data. Two main types of data were generated by our standardized behavioural assay. First, we used trials to establish whether or not birds used tools, irrespective of deployment context and extraction success (see Fig. 2b, f). Second, for those birds that did use tools, we examined aspects of tool handling, modification (and possible manufacture) and deployment, and quantified the speed with which they extracted bait from the log's holes and crevices (see Fig. 2g; trials included only when birds had been tested individually). Formal species comparisons are pending, but when extracting meat from vertical holes, 'Alalā's performance ($n = 52$ birds that probed; 63% of attempted extractions successful; cumulative probing time until extraction (median, range): 26.8 s, 3.2–215.6 s; see top-left panel of Fig. 2g) is broadly comparable to that of New Caledonian crows (more difficult, deeper and narrower holes³: $n = 15$; 49%; 42.3 s, 5.8–161.6 s; unpublished data).

Visual-field measurements require that subjects' heads are held completely still for ~ 30 – 45 min⁹. While such temporary restraint is tolerated well by most birds, it cannot currently be used with 'Alalā, given the species' critical conservation status. As the width of the binocular field is determined to a large degree by lateral eye-movement amplitude (correlation, $r = 0.82$, $P = 0.02$, $n = 7$ *Corvus* spp.; data from table 1 in ref. 9), we opportunistically assessed, during behavioural trials and when handling subjects for routine health checks, how much birds can rotate their eyes forward during full convergence (see Extended Data Fig. 1b and Supplementary Video 5).

Ontogenetic patterns. To gain insights into possible genetic predispositions^{15,16,36}, we studied the development of object-oriented behaviour in seven juvenile 'Alalā that had been bred and puppet-reared³⁷ at the KBCC in 2012 (hatch dates between 20 June and 16 July). Subjects were housed in two mixed-parentage groups (offspring of five different pairs) of three (Group A: subjects #206, #207, #208) and four birds (Group B: #200, #201, #204, #205), respectively. Following the facility's standard procedures, birds were transferred from fledgling aviaries ($\sim 2.0 \times 1.8 \times 2.3$ m) to large outdoor aviaries after they had acquired basic flight skills, at 61–69 days old. From 15 September onwards, the groups were housed in adjacent aviary chambers (each $\sim 3.0 \times 12.0 \times 5.5$ m), with visual contact through a wire-mesh partition, but they never saw adults during the full duration of our study. Furthermore, all staff were briefed never to use 'tools' (of any kind) in front of subjects, both during formal observation sessions and in all other contexts, including general husbandry activities (owing to an oversight, large metal tongs were used on a few occasions, to scrape old food from logs). As subjects were co-housed in groups, individuals that only expressed tool use later in the experiment could potentially have learned from those that used tools earlier (see Fig. 3c). This means that only the very first tool behaviour expressed in either of the two experimental groups was certain to be an independent 'discovery'^{15,19}.

We collected two main data sets. First, we employed a standard focal-bird observation protocol^{15–17} to document the natural development of object-oriented

behaviour. Up to three days per week (usually on Tuesday, Thursday and Saturday), we conducted a morning (between ~6:30–11:00 h) and an afternoon (~12:00–16:00 h) session, aiming to collect ~5 min of video footage per subject (that is, 3×2 sessions \times 5 min = 30 min, per subject per week). To avoid biases, the order in which groups were observed, and the order in which subjects were observed within sessions, was pseudo-randomized, and session start times were varied slightly within the above-mentioned time windows. Second, once per week (usually on Fridays), we conducted a 'probe trial' to assess subjects' tool-use competence. We presented each group for ~15–20 min with a wooden platform, containing food-baited vertical holes and crevices (Extended Data Fig. 2b). The rationale of our study design was to monitor the development of the subjects' tool-related behaviour (see Fig. 3c) with minimal environmental 'scaffolding'; note that, in contrast, the New Caledonian crows raised in an earlier study had *ad libitum* access to extraction tasks^{15,16}.

Platforms were initially baited with waxworms and cereal treats, but from 5 October 2012 onwards, we switched to mouse heads, neonate mice, and bright-red 'Ohelo *Vaccinium reticulatum* berries³⁸. By January 2013, subjects in both groups showed keen interest in the hidden food, and often handled objects near the platform. For two reasons, however, their tool-use attempts largely failed: they sourced inappropriate materials as tools (for example, decaying pieces of fern), and even when suitable sticks were found, they struggled to extract food from tasks. We addressed these problems by providing sticks of assorted length (6 of 10–15 cm; 6 of 20–25 cm), loosely placed in the centre of the platform (sticks were never handled in view of the birds, and never pre-inserted into tasks), and by adding horizontal holes and crevices from which food was presumably easier to extract. These changes implemented, we concluded our experiment by providing birds with abundant opportunities to practice their tool-use skills (see entries ①–③ in Fig. 3c; trial length extended to ~30 min), with: a week of almost daily platform trials (23–29 January 2013; pooled data shown as ①); two re-test trials about a week later (4 and 6 February 2013; pooled data shown as ②); and another 1.5 weeks of exposure to the platform and a range of other extraction tasks without observation (8–18 February 2013), followed by a final platform trial on 20 February 2013 (entry ③). For reference, when protocols were altered on 23 January 2013, subjects were 151–181 days post fledging.

Following standard protocols, subjects received near-daily aviary enrichment (sometimes immediately prior to observation sessions), including a variety of food items that required processing but were accessible by bill alone. The exception to this was baited opaque PVC tubes, which were presented on a single day in weeks 11, 12, 16, 19 and 24 (with week 1 commencing on 3 September 2012), to assess how birds' tool-related performance on this task compared to that expressed during formal probe trials with the more demanding platform-mounted set-up (see above). These sessions were not included in focal-bird analyses shown in Fig. 3a, but some object insertions were documented slightly ahead of formal platform probe trials (Fig. 3c).

Videos from all observation sessions were scored with JWatcher software³⁹ in randomized order by two hypothesis-naïve observers (S.W. and Caitlin Higgott), who achieved very high inter-observer agreement for a subsample of three sessions (correlation coefficients for handling rates for the object categories shown in Fig. 3a, $r = 0.96$ – 0.99 , all $P < 0.0001$, $n = 10$ scores for each test); sessions for post-fledging weeks 3–5 (data from fledgling aviaries included) were scored with a particularly detailed scheme, with some behaviours coded as states, rather than as events, for time-budget analyses (weekly sample sizes were 3, 5 and 7 birds, respectively; Fig. 3d). We wrote code in R⁴⁰ for extracting data from raw JWatcher output files, to calculate either object-handling rates (Fig. 3a; data for 'sticks' and 'stones' analysed with simple correlations) or time budgets (Fig. 3d; calculated for the time focal subjects were in view). Except for cross-species comparisons (see below), we plotted temporal data by calendar week (Fig. 3a, c), rather than by bird age or time since fledging, because the development of the younger birds in Group A may have been accelerated through observing the older members of Group B in the adjacent aviary chamber. In videos of probe trials, we scored which behavioural actions subjects performed near or on the platform, ranging from merely approaching the set-up to successfully using tools to extract bait (action types are numbered in the panels of Fig. 3c, and descriptions are provided in Extended Data Table 1).

For cross-species comparisons, we extracted data on the development of object-oriented behaviour in New Caledonian crows and common ravens from figure 2 in ref. 17. For 'stick' manipulation, we only used data from untutored New Caledonian crows (2 subjects)¹⁷, and the object category 'perch' included all non-portable aviary fixtures. These species comparisons are for indicative purposes only (Fig. 3d), as the three studies considered varied in a range of factors, including details of subject housing, access to objects and extraction tasks, observation conditions and behavioural scoring (note considerable variation for 'stick'

estimates for 'Alalā'), and the species in question are known to exhibit different rates of juvenile development^{4,5,8}.

Historical observations. Prior to the commencement of our study, 'Alalā had regularly been observed using tools in both captive facilities. Staff did not consider these cases particularly noteworthy, as they were aware that the behaviour had been previously described for the congeneric New Caledonian crow. To provide context for our study, we collated information on these earlier, opportunistic observations, trying to locate written records^{12,13} and conclusive photo or video evidence (Supplementary Video 4). It is worth noting that our sample of well-documented historical observations constitutes only a small fraction of the observations made by facility staff over the years.

Correlates of phenotypic variation. To examine the influence of environmental and/or social factors on tool-use competence, we reconstructed our subjects' life-time housing histories—that is, the time they had spent at different facilities, their allocation to particular aviaries and chambers, and their co-housing with other birds—using paper files and electronic spreadsheets held at the KBCC and MBCC.

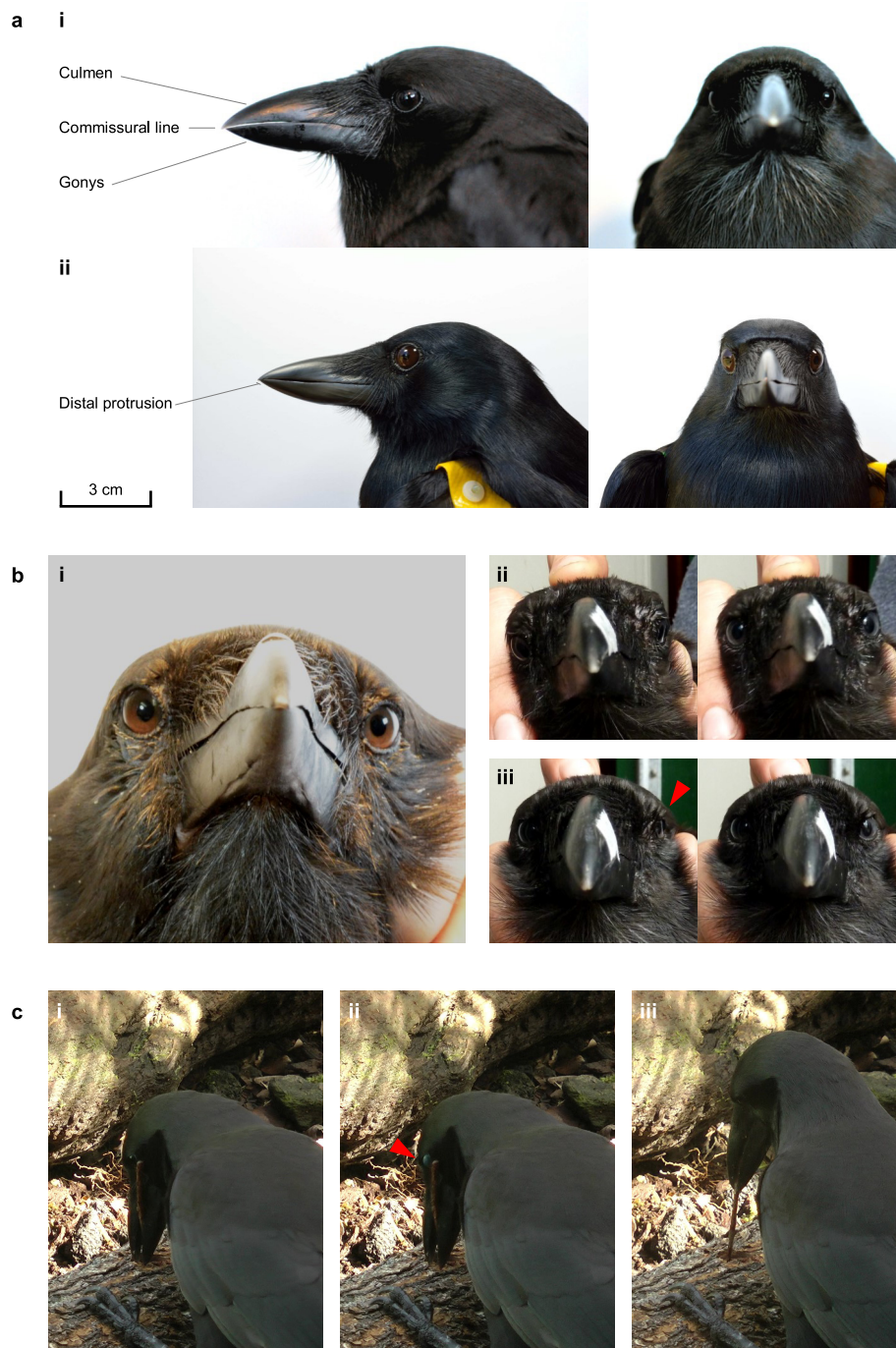
First, we conducted some basic checks, to see whether competence was related to being raised (first two years of life), or kept, in a particular facility (Fig. 2c). Next, we used our detailed housing data to investigate how well our study population was admixed socially, by simulating⁴¹ the flow of information—such as tool use—across birds^{42,43}. Using all dated housing entries in our database ($n = 1,501$ for 135 birds in 1996–2013), we first generated contact networks that specified which crow dyads were in potential visual contact at any given time, by sharing an aviary or occupying adjacent aviaries/chambers with a see-through wire-mesh partition (cumulative 'co-housing matrix' shown in Fig. 3e, left). As the expression of 'Alalā tool behaviour is strongly age-dependent (Fig. 2d), and studies in other systems have shown that learning is often particularly effective during a 'sensitive window' early in life¹⁸, we considered only the subset of co-housing events in which one of the birds was adult (>2-years-old) and the other immature (<2-years-old). Our idealised simulation model assumed that, if the adult had the information at the time of co-housing, it was expressed and transmitted instantaneously to the immature. The information was never lost, so both the adult (and the immature, once old enough) could pass it on in subsequent co-housing events. We then traced (computationally) for all potential 'innovators' of information all possible transmission pathways through the time-ordered contact networks, identifying those reaching confirmed tool users by 2013 (grey dots in Fig. 3e, left, refer to immature recipients that were not among the confirmed tool users in 2013); the results are summarized in the 'reachability matrix' (Fig. 3e, right). From this matrix we computed⁴⁴ the smallest number (m) of independent innovation events (rows) needed to ensure that every tool user (column) is reached. For the transmission dynamics described, $m = 8$. To establish a lower-bound estimate, we relaxed the transmission rules so that information could be passed between birds of all ages, yielding $m = 1$. Both simulations assumed highly conservatively that transmission was not only instantaneous but also deterministic (although we would expect considerable between-dyad variation in transmission probabilities due to differences in social-learning opportunities and phenotypic plasticity^{18,45}), but inevitably had to ignore possible pathways created by birds for which exact aviary information was unknown (16.3% of 1,501 housing entries). As explained in the main text, these analyses helped us to characterize the 'social connectivity' of our study population, but further behavioural experiments are required to demonstrate social learning in 'Alalā.

Phylogenetic relationships. To examine phylogenetic relationships within the genus *Corvus*, we built a consensus tree (Fig. 1a) from sequence data that had previously been archived in GenBank by two independent studies^{6,46} (note that *C. macrorhynchos culminatus* had erroneously been logged as *C. culminatus* in GenBank⁶). Where more than one sequence was available for a given species, we aligned them and produced a consensus sequence. We then aligned each region (CR, GAPDH, ND2, ND3, and ODC) separately using MAFFT⁴⁷, and concatenated these alignments. For species that did not have coverage for a particular region, these regions were coded as Ns. We used this alignment to generate a consensus tree, using MrBayes⁴⁸ ($n_{\text{gen}} = 10,000,000$). Uncertainty about the specific status of some taxa affects the total number of species within the genus^{6,8,46} (for example, recent authors⁴⁶ treated *C. violaceus* and *C. minutus* as distinct species, rather than as subspecies of, respectively, *C. enca* and *C. palmarum*⁸), but not the gross topology of the phylogenetic tree. Importantly, although more work is required to resolve the close relationships of *C. moneduloides*^{4,6,46}, our analyses confirmed that the two tool-using species *C. hawaiiensis* and *C. moneduloides* are only very distantly related⁴⁹. While our concatenation method enabled us to maximise data coverage, it complicated the estimation of divergence times; according to an earlier study, however, the last common ancestor would have lived in the mid-Miocene, ~11 million years ago (see figure 2 in ref. 46).

The 'Alalā is the only survivor of at least five species of crow that once inhabited the Hawaiian archipelago^{5,26,30}. To assess variation in craniofacial features, we used

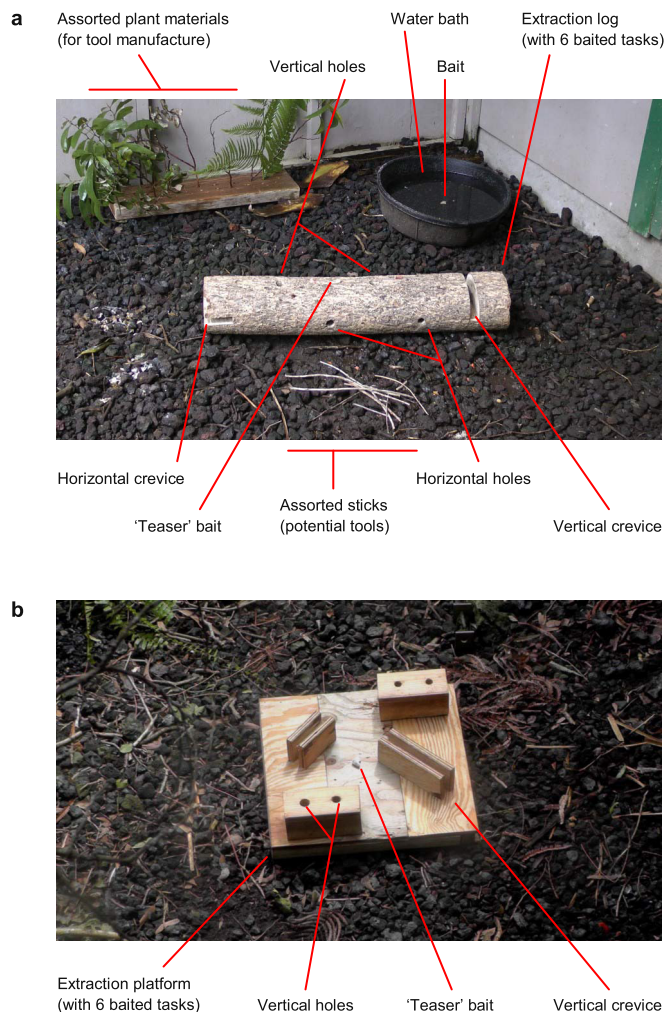
previously published photos (figure 3 in ref. 26) of the fossil skulls of two extinct species (*C. impluviatus*, *C. viriosus*), and adapted (mandibles closed; flipped horizontally; re-coloured) and re-sized them for direct comparison with the portrait photo of a live 'Alalā (adult female #94; Fig. 1f). The evolutionary history of this species assemblage remains unknown, but variation in bill morphology indicates well-differentiated foraging behaviour^{50,51}. The distribution of an undescribed species with “a bill modified for hammering”⁵ may be of particular relevance⁴ for understanding the evolutionary ecology of tool behaviour in 'Alalā.

30. Banko, P. C. in *Conservation Biology of Hawaiian Forest Birds* (eds Pratt, T. K., Atkinson, C. T., Banko, P. C., Jacobi, J. D. & Woodworth, B. L.) (Yale Univ. Press, 2009).
31. National Research Council (NRC). *The Scientific Bases for the Preservation of the Hawaiian Crow* (National Academy Press, 1992).
32. Hedrick, P. W., Hoeck, P. E. A., Fleischer, R. C., Farabaugh, S. & Masuda, B. M. The influence of captive breeding management on founder representation and inbreeding in the 'Alalā, the Hawaiian crow. *Conserv. Genet.* **17**, 369–378 (2016).
33. Griffiths, R., Double, M. C., Orr, K. & Dawson, R. J. G. A DNA test to sex most birds. *Mol. Ecol.* **7**, 1071–1075 (1998).
34. Harvey, N. N., Farabaugh, S. M. & Druker, B. B. Effects of early rearing experience on adult behavior and nesting in captive Hawaiian crows (*Corvus hawaiiensis*). *Zoo Biol.* **21**, 59–75 (2002).
35. Péter, A. Solomon Coder (version beta14.03.10): A Simple Solution for Behavior Coding. <http://solomoncoder.com> (2011).
36. Mamelì, M. & Bateson, P. An evaluation of the concept of innateness. *Phil. Trans. R. Soc. B* **366**, 436–443 (2011).
37. Kuehler, C., Kuhn, M., McIlraith, B. & Campbell, G. Artificial incubation and hand-rearing of 'Alalā (*Corvus hawaiiensis*) eggs removed from the wild. *Zoo Biol.* **13**, 257–266 (1994).
38. Culliney, S., Pejchar, L., Switzer, R. & Ruiz-Gutierrez, V. Seed dispersal by a captive corvid: the role of the 'Alalā (*Corvus hawaiiensis*) in shaping Hawai'i's plant communities. *Ecol. Appl.* **22**, 1718–1732 (2012).
39. Blumstein, D. T. & Daniel, J. C. *Quantifying Behavior the JWatcher Way* (Sinauer Associates, 2007).
40. R Development Core Team. R: A Language and Environment for Statistical Computing. <http://www.R-project.org> (2014).
41. St Clair, J. J. H. et al. Experimental resource pulses influence social-network dynamics and the potential for information flow in tool-using crows. *Nat. Commun.* **6**, 7197 (2015).
42. Auersperg, A. M. et al. Social transmission of tool use and tool manufacture in Goffin cockatoos (*Cacatua goffini*). *Proc. R. Soc. B* **281**, 20140972 (2014).
43. Hobaiter, C., Poisot, T., Zuberbühler, K., Hoppitt, W. & Gruber, T. Social network analysis shows direct evidence for social transmission of tool use in wild chimpanzees. *PLoS Biol.* **12**, e1001960 (2014).
44. Gori, F., Folino, G., Jetten, M. S. M. & Marchiori, E. MTR: taxonomic annotation of short metagenomic reads using clustering at multiple taxonomic ranks. *Bioinformatics* **27**, 196–203 (2011).
45. Hoppitt, W. & Laland, K. N. *Social Learning: An Introduction to Mechanisms, Methods and Models* (Princeton Univ. Press, 2013).
46. Jönsson, K. A., Fabre, P.-H. & Irestedt, M. Brains, tools, innovation and biogeography in crows and ravens. *BMC Evol. Biol.* **12**, 72 (2012).
47. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
48. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).
49. Fleischer, R. C. & McIntosh, C. E. Molecular systematics and biogeography of the Hawaiian avifauna. *Stud. Avian Biol.* **22**, 51–60 (2001).
50. Gibbs, H. L. & Grant, P. R. Oscillating selection on Darwin's finches. *Nature* **327**, 511–513 (1987).
51. Lamichhaney, S. et al. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518**, 371–375 (2015).
52. Tebbich, S., Taborsky, M., Fessl, B. & Dvorak, M. The ecology of tool-use in the woodpecker finch (*Cactospiza pallida*). *Ecol. Lett.* **5**, 656–664 (2002).
53. Morse, D. H. The use of tools by brown-headed nuthatches. *Wilson Bull.* **80**, 220–224 (1968).
54. Visalberghi, E. & Fragaszy, D. in *Tool Use in Animals: Cognition and Ecology* (eds Sanz, C. M., Call, J. & Boesch, C.) (Cambridge Univ. Press, 2013).



Extended Data Figure 1 | Craniofacial morphology of tool-using ‘Alalā and New Caledonian crows. **a**, Although some other *Corvus* species⁸ have relatively straight bills—in terms of culmen- and commissural-line projections—they usually lack the pronounced distal angle of the gonys that is characteristic of the tool-using (i) ‘Alalā (adult female #191, 8 January 2015) and (ii) New Caledonian crow (adult female #CR6, 6 October 2013; photo: P. Barros da Costa), and also have larger distal protrusions of the upper mandible. Despite the overall similarity of the two species^{4,10}, ‘Alalā are larger and more robust birds (Fig. 1c, d), and exhibit modest bill curvature, relatively smaller eyes, and notable intraspecific variation in bill shape. The scale bar applies to all four images. **b**, ‘Alalā have markedly forward-pointing eyes, with high lateral eye-movement amplitudes, enabling (i) a considerable degree of convergence (#96, 17 February 2014; note that the red-brown plumage colouration is an image artefact; no adjustments have been made). The movement of (ii) both eyes

(#201, 9 August 2014), or (iii) just one eye (red arrow; #206, 9 August 2014), can often be observed during the handling of birds for routine health checks (the white marker on the bills is a removable scale bar; see Supplementary Video 5). Although the ‘Alalā’s visual field could not be measured in this study (see Methods), these features are likely to produce a large field of binocular overlap, which in New Caledonian crows is thought to aid tool manufacture and deployment⁹. **c**, When ‘Alalā hold stick tools in a transverse grip, (i) the slight curvature of the birds’ bill can force the non-functional end of the tool close to the eye (as would be predicted from earlier work; see figure 5 in ref. 9), (ii) which may cause discomfort or even injury (red arrow indicates nictitating membrane, which the bird closed temporarily to protect its eye); (iii) this may explain why the vast majority of individuals prefer to hold tools in a frontal grip (adult male #134, 21 January 2013; transverse grip observed in only 11 of 104 subjects tested on the standardized log task).



Extended Data Figure 2 | Food-extraction tasks for investigating tool-use behaviour in captive 'Alalā. **a**, A species-wide assay of tool-use competence was conducted by presenting birds with a baited *Koa Acacia koa* log (length, ~78 cm; diameter, ~16 cm), containing two vertical holes (depth, ~5.0 cm; diameter, ~2.3 cm), two horizontal holes (~5.4 cm; ~2.3 cm), one vertical crevice (width × depth, ~2.4 × 6.4 cm) and one horizontal crevice (height × depth, ~2.3 × 6.8 cm); all estimates of dimensions are averages for the two log set-ups used in experimental trials (see Methods). Sticks for potential tool use were scattered in front of the log (length classes: 3 of 0–5 cm; 3 of 10–15 cm; 3 of 20–25 cm; and 3 of 30–35 cm), and assorted plant materials for potential tool manufacture were provided on a wooden stand nearby (KBCC: 2 'Ōhi'a lehua *Metrosideros polymorpha* stems, 2 *Koa* stems, 1 fern frond, 2 dead branched stems; MBCC: 2 dead branched stems). As subjects had access to suitable tools during trials, current data probably underestimate the species' tool-making capabilities. **b**, The tool-use competence of seven juvenile birds was assessed once per week (and more often towards the end of the study period; see Methods), using a baited wooden platform (~50 × 50 cm) with four vertical holes (depth, ~4.5–5.4 cm; diameter, ~2.0–2.7 cm) and two vertical crevices (width × length × height, ~2.5 × 21.2 × 7.3 cm and ~2.4 × 13.5 × 8.0 cm). From late January 2013 onwards, a second replica platform was used to enable parallel testing of both experimental groups. During the final stages of the experiment, the four vertical holes were substituted with horizontal holes (by rotating the wooden blocks), and two horizontal crevices were added (not shown here; see Supplementary Video 3).

Extended Data Table 1 | Behavioural actions scored for captive, juvenile 'Alalā during standardized probe trials

Action type	Description
18*	Tool-assisted bait extraction from horizontal crevice (insert → move → acquire) [†]
17	Tool-assisted bait extraction from vertical crevice (insert → move → acquire) [†]
16*	Tool-assisted bait extraction from horizontal crevice (insert → move) [†]
15	Tool-assisted bait extraction from vertical crevice (insert → move) [†]
14*	Tool-assisted bait extraction from horizontal crevice (move → acquire) [†]
13	Tool-assisted bait extraction from vertical crevice (move → acquire) [†]
12	Inserting bill-held other natural object [§] into hole or crevice
11	Inserting bill-held stick-type object [‡] into hole or crevice
10	Combining bill-held other natural object [§] within the platform area [¶]
9	Combining bill-held stick-type object [‡] within the platform area [¶]
8	Dropping other natural object [§] (picked up within the platform area [¶]), within the platform area [¶]
7*	Dropping provided stick [‡] within the platform area [¶]
6	Dropping non-provided stick-type object [‡] (picked up within the platform area [¶]), within the platform area [¶]
5*	Extraction from horizontal crevice without tool [#]
4	Extraction from vertical crevice without tool [#]
3	Extraction from hole without tool [#]
2	Chiselling at hole or crevice
1	Within the platform area [¶]

Action types correspond to the numbers shown on the y-axes of panels in Fig. 3c; for a photograph of the baited experimental platform, see Extended Data Fig. 2b. Action types are grouped into: approach to and interaction with the platform, not directly involving objects (no shading); object dropping near or on the platform (grey); object combinations and insertions (includes unsuccessful tool use) near or on the platform (light blue); and successful bait extractions with tools from platform tasks (dark blue).

*These actions were possible only after the platform set-up had been modified during the final stages of the experiment (see Methods).

†'Tool-assisted bait extractions' were scored according to which parts of an extraction sequence a bird performed. Full sequences involved tool insertion, movement of the tool, and bait acquisition (insert → move → acquire), but in some cases, birds failed to acquire the bait they had brought within reach because another subject took it (insert → move), or they used a tool for extraction that had previously been inserted by another subject (move → acquire). Bait extractions were often highly dynamic, involving multiple birds (see Scene B of Supplementary Video 3). Note that there were no tool-assisted bait extractions from holes.

‡'Stick-type object' is an elongated object that could potentially be used as a tool; some side branches may remain, but the object must have a potentially functional end that could be inserted into holes or crevices (some sticks were provided on the platform during the final stages of the experiment; see Methods). Note that, for the purpose of estimating object-handling rates and manipulation times for Fig. 3a and 3d, respectively, 'sticks' additionally include fern sections and branched pieces of plant for 'Alalā (see objects lying around the platform in Scene B of Supplementary Video 3), given that: these were common objects in the housing aviaries; birds inserted them into baited platform tasks; and another study used a similarly inclusive object class 'twigs' for other corvid species¹⁷.

§'Other natural object' includes fern sections, branched pieces of plant, larger branches, leaves, grass blades, bark pieces, wooden splinters and stones (but not food objects) that do not fulfil the definition of 'stick-type object'. Note that, in Fig. 3a, 'other objects' also include toys and other enrichment items, and 'stones' are shown separately.

||'Combining' is any manipulation in which an object is placed in contact with another object or substrate, but not including 'insertions' (compare with ref. 17). Combinations include attempted insertions with stick-type objects oriented at the wrong angle, as well as the placement and movement of objects.

¶'Within the platform area' was scored when a bird was on the platform itself, or its bill tip or any object held in its bill was within ~10 cm of the platform (~one quarter of a bird's body length).

#While cases were omitted where a bird's bait acquisition was immediately preceded by tool movements by another subject, it is possible that tool-bait interactions earlier during the trial were responsible for bringing bait within bill reach.

Extended Data Table 2 | Observation rates of tool behaviour for three naturally tool-using bird species

Study species	Habitat (conditions)	Observation time (h)	Tool-use observations	Tool-use observations h ⁻¹
New Caledonian crow	Coastal dry forest	9.2	8	0.9
Woodpecker finch	Humid <i>Scalesia</i> zone	7.2	6	0.8
	Arid zone	14.1	134	9.5
Brown-headed nuthatch	Pine forest (few seeds)	150	10	0.07
	Pine forest (abundant seeds)	75	1	0.01

The most detailed study on the foraging behaviour of free-ranging 'Alalā accumulated ~17.5 h of focal observations for eight pairs in montane rainforest²¹, and although a sample like this would almost certainly yield conclusive tool-use observations in some habitual avian tool users (New Caledonian crow²³; woodpecker finch⁵²), it would not necessarily be sufficient for others (brown-headed nuthatch⁵³). For comparison, orang-utans *Pongo* spp. and capuchin monkeys *Cebus/Sapajus* spp. were long thought to use tools exclusively in captivity, and it took decades of high-effort fieldwork to uncover the diverse tool behaviours of wild populations^{1,54}.

Life history of the stem tetrapod *Acanthostega* revealed by synchrotron microtomography

Sophie Sanchez^{1,2}, Paul Tafforeau², Jennifer A. Clack³ & Per E. Ahlberg¹

The transition from fish to tetrapod was arguably the most radical series of adaptive shifts in vertebrate evolutionary history. Data are accumulating rapidly for most aspects of these events^{1–5}, but the life histories of the earliest tetrapods remain completely unknown, leaving a major gap in our understanding of these organisms as living animals. Symptomatic of this problem is the unspoken assumption that the largest known Devonian tetrapod fossils represent adult individuals. Here we present the first, to our knowledge, life history data for a Devonian tetrapod, from the *Acanthostega* mass-death deposit of Stensjö Bjerg, East Greenland^{6,7}. Using propagation phase-contrast synchrotron microtomography (PPC-SRμCT)⁸ to visualize the histology of humeri (upper arm bones) and infer their growth histories, we show that even the largest individuals from this deposit are juveniles. A long early juvenile stage with unossified limb bones, during which individuals grew to almost final size, was followed by a slow-growing late juvenile stage with ossified limbs that lasted for at least six years in some individuals. The late onset of limb ossification suggests that the juveniles were exclusively aquatic, and the predominance of juveniles in the sample suggests segregated distributions of juveniles and adults at least at certain times. The absolute size at which limb ossification began differs greatly between individuals, suggesting the possibility of sexual dimorphism, adaptive strategies or competition-related size variation.

The life cycle of the earliest tetrapods, and its role in the transition from water to land, has long been a matter of speculation. For example, it has been suggested that the earliest tetrapods bred in ephemeral pools, and that the need for the larvae to locomote overland or through extremely shallow water when relocating from these drying ponds to more permanent water bodies provided selective pressure towards the evolution of terrestriality⁹. However, the fossil record of Devonian tetrapods, being dominated by rare and incomplete specimens that frequently come from poorly constrained localities such as scree slopes¹⁰, has until now yielded virtually no life history data.

The only known Devonian tetrapod locality with good potential for revealing life history information is the *Acanthostega* mass-death deposit in the Britta Dal Formation (Upper Devonian, Famennian) on Stensjö Bjerg, East Greenland¹⁰. This locality, comprising a small *in situ* micaceous silty sandstone body and immediately associated scree¹⁰, has yielded more than 200 skeletal elements. Fourteen skulls, six of them associated with partially articulated skeletons, were complete enough to measure¹¹, and several more can be identified as individuals; there must have been at least 20 animals represented, although almost certainly more were present. Other vertebrates are represented only by a few isolated bones⁷. The *Acanthostega* individuals in this deposit evidently died together, probably during drought following a sheet-flood event⁶; they thus represent a single time-point sample from a population of this stem tetrapod.

Acanthostega humeri from the mass-death deposit show varying degrees of ossification, representing a possible partial ontogenetic

series^{11,12}. Using the non-destructive imaging technique PPC-SRμCT⁸, performed at beamline ID19 of the European Synchrotron Radiation Facility (ESRF) (see Methods for more details), we have undertaken histological investigations of the four humeri collected from the locality (Natural History Museum of Denmark MGUH 29019, MGUH 29020, NHMD 74756; University Museum of Zoology Cambridge UMZC T.1295)¹³ (Fig. 1a and Extended Data Fig. 1), recovering data that illuminate the life history of *Acanthostega*. These are all the humeri of *Acanthostega* known so far. The humerus MGUH 29019 comes from an articulated specimen. The other humeri are isolated bones. The humeri fall into two distinct size classes—large (NHMD 74756, MGUH 29020) and small (MGUH 29019, UMZC T.1295) (Fig. 1a and Extended Data Fig. 1). Consistent with previous observations¹², we found no correlation between size and degree of ossification: specimens NHMD 74756 and UMZC T.1295 are weakly ossified whereas specimens MGUH 29019 and MGUH 29020 are strongly ossified (Fig. 1a and Extended Data Fig. 1).

All humeri exhibit an extensive spongiosa surrounded by a thin compact cortex (Fig. 1a, b). This arrangement resembles that of the humerus of the lobe-finned fish *Eusthenopteron*¹⁴, a less crownward member of the tetrapod stem group¹⁵. Remnants of calcified cartilage in the metaphyseal region (close to the articular extremities; Extended Data Fig. 2c) show that the spongiosa formed by endochondral ossification as in extant tetrapods¹⁶ and *Eusthenopteron*^{14,17}. Tubular structures at the base of the epiphyses (Extended Data Fig. 2a, b) resemble the marrow processes in the growth plate of the humerus of *Eusthenopteron*¹⁴.

The midshaft cortex of all *Acanthostega* humeri contains a dense arrangement of radial vascular canals (Fig. 1c) similar to that of juvenile *Eusthenopteron*¹⁴. The radial canals connect to a basal mesh of surface-parallel canals (Fig. 1c). In the largest specimen, MGUH 29020, the radial canals vary in diameter between different parts of the scanned area (Fig. 1c), probably reflecting local blood-supply needs. Although the cortex shows evidence of patchy basal erosion in three of the humeri, all appear to retain areas of primary internal cortical surface (Fig. 2). Clusters of large aligned globular cell lacunae between the endosteal bone and the cortex (Fig. 2b, d and Extended Data Fig. 3) can be identified as chondrocyte lacunae by comparison with juvenile *Eusthenopteron*, in which similar lacunae lie between the cortical bone and unresorbed remnants of calcified cartilage¹⁴. These numerous alignments of chondrocyte lacunae at midshaft (Fig. 2 and Extended Data Fig. 3) mark the perichondral surface of the original cartilaginous humerus. As limb bone growth originates at the midshaft, this implies that the cartilaginous rod was very large relative to the observed final size of the bone, and that cortical bone growth conversely made only a modest contribution to the final size. In other words, the *Acanthostega* individuals grew almost to full observed size before their humeri began to ossify.

¹Science for Life Laboratory and Uppsala University, Subdepartment of Evolution and Development, Department of Organismal Biology, Evolutionary Biology Centre, Norbyvägen 18A, 752 36 Uppsala, Sweden. ²European Synchrotron Radiation Facility, 71 Avenue des Martyrs, CS-40220, 38043 Grenoble Cedex, France. ³University Museum of Zoology, Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK.

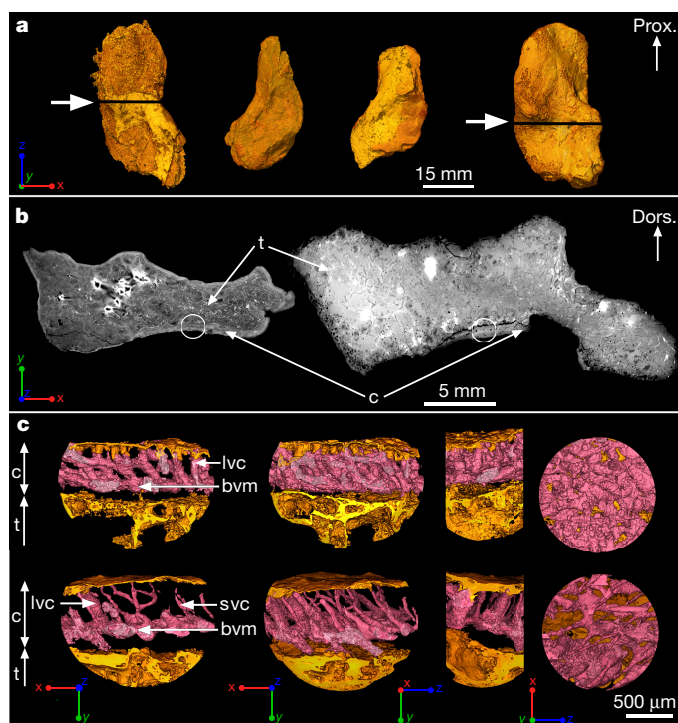


Figure 1 | Midshaft bone microanatomy and histology of *Acanthostega* humeri. **a**, Three-dimensional models of humeri shown in ventral view, based on synchrotron data. From left to right: NHMD 74756 (voxel size: 14.95 µm), UMZC T.1295 (voxel size: 12.62 µm), MGUH 29019 (voxel size: 12.62 µm) and MGUH 29020 (bottom, voxel size: 20.24 µm). The white arrows indicate the locations of the virtual thin sections in **b**. **b**, Humeral microanatomy of NHMD 74756 (left) and MGUH 29020 (right) showing an extended trabecular cavity (t) surrounded by a thin layer of compact cortical bone (c). The transverse virtual thin sections are 80 µm thick. The white circles indicate the locations of the high-resolution scans modelled in **c**. **c**, Three-dimensional models (voxel size: 0.638 µm) of the cortical bone microstructure of NHMD 74756 (top) and MGUH 29020 (bottom) showing a dense oblique mesh of large (lvc) and small (svc) vascular canals (in pink) connected to a horizontal basal vascular mesh (bvm). From left to right: transverse 3D thin section (250 µm thick), complete 3D model in transverse orientation, 3D model in longitudinal orientation and tangential section showing the inner view of the 3D vascular mesh. Dors., dorsal; prox., proximal.

The presence of lines of arrested growth (LAGs) in the cortical bone permits us to infer how many years were occupied by the deposition of this tissue, on the assumption that the deposit between two LAGs represents an annual cycle, as in most extant tetrapods^{18,19}. Observations in homologous regions (Extended Data Fig. 4) of the four humeri reveal a maximal number of six LAGs in MGUH 29020 (Fig. 3b, d and Extended Data Fig. 5b, e), four in NHMD 74756 (Fig. 3c, e) and UMZC T.1295 (Extended Data Fig. 5d), and three in MGUH 29019 (Extended Data Fig. 5a, c). All observations were made in areas that were at least partly covered with matrix and thus unlikely to have been affected by external erosion. These LAG patterns are regular and show no tightening (Extended Data Table 1)—that is, no deceleration of the growth rate (Extended Data Fig. 6)—as would be expected at sexual maturity in adult tetrapods^{19–21}. This suggests that the four specimens of *Acanthostega* were still juveniles when they died, assuming that their humeri had begun to ossify before the onset of sexual maturity (as they do in all known tetrapods^{22–24} and in *Eusthenopteron*¹⁴; see Supplementary Information). The juvenile stage must therefore have lasted at least six years in *Acanthostega*. Indeed, it probably lasted a good deal longer, because the cartilaginous humerus grew to almost full size before cortical bone deposition, and thus the recording of annual growth increments, even began. *Acanthostega* is not the only member of the tetrapod stem group to show late onset of ossification.

Juvenile *Eusthenopteron* exhibits a large spongiosa and a cortex with no internal resorption, showing that the original cartilaginous rod was approximately two-thirds of adult spongiosa size and presumably formed over several years¹⁴. How this relates to final adult size in *Acanthostega* is difficult to say, but the slow growth rate of the juvenile *Acanthostega* suggests that final adult size may not have been much greater than the largest individuals recorded from the mass-death deposit.

The complete lack of correlation between size and degree of ossification could reflect some form of individual variation, such as the competition-related size variation observed in certain extant tetrapods²⁵, adaptive strategies or sexual dimorphism²⁶. Under these interpretations, some individuals (represented by MGUH 29019 and UMZC T.1295) began to ossify their humeri—and presumably approach sexual maturity—at a much smaller size than others (represented by MGUH 29020 and NHMD 74756). Unfortunately, the very small sample size does not allow us to determine whether the apparently discrete size classes reflect a real bimodal size distribution, or whether they are simply the outcome of randomly sampling a continuous size variation. However, the observed combination of sizes and ossification states categorically invalidates the construction of an ontogenetic sequence from smallest to largest humerus.

The synchrotron virtual histological data from the humeri shed new light on several aspects of the palaeobiology and life history of *Acanthostega*. It had a prolonged juvenile stage, no less than six years (as shown by the LAGs) but more probably at least a decade, given that it grew almost to full recorded size before the onset of cortical bone ossification. This aligns it with a range of sarcopterygian fishes and tetrapods including *Neoceratodus* (15–20 years²⁷), *Eusthenopteron* (adulthood at 11 years¹⁴), *Discosaurus* (10 years²⁰) and *Andrias* (larval period of 4–5 years and 10 years to adulthood²⁸), suggesting that a long juvenile stage could be primitive for tetrapods. The late onset of ossification in *Acanthostega* implies that the early juvenile stage was aquatic, as a cartilaginous humerus would be ill-suited for terrestrial locomotion; this also agrees with the presence of aquatic adaptations such as a large caudal fin and well-developed gill skeleton in *Acanthostega*^{1,29}, and contradicts the hypothesis of juvenile terrestriality⁹ at least for this particular tetrapod.

The fact that all four humeri appear to belong to juvenile individuals suggests that the mass-death assemblage is dominated by, and may in fact consist exclusively of, juveniles. The assemblage does not include a subset of distinctively larger individuals. Specimens MGUH 29019 and MGUH 29020 are the most fully ossified humeri¹² of the assemblage. MGUH 29019, the smallest humerus, is associated with a 12-cm-long skull; MGUH 29020, the largest humerus, is an isolated find from the scree but appears to represent one of the largest individuals in the assemblage (personal observation, J.A.C. and P.E.A.).

The palaeoenvironmental data from the locality provide a context for these observations. It forms part of a large ephemeral fluvial system in an otherwise arid tropical landscape⁶, extending northwards for more than 200 km from an unpreserved source water body that must have been large and permanent as it housed large lobe-finned fishes such as *Eusthenodon* and *Holoptychius*⁷. The *Acanthostega* individuals appear to have been flushed out into this fluvial system during a flood event, after which the ensuing drought concentrated them in a shrinking pool and eventually killed them⁶. The almost complete absence of other taxa in the death assemblage suggests that it is not a concentrate of a whole fauna (like the near-contemporary mass-death deposit from Canowindra, Australia³⁰) but rather a reflection of schooling behaviour in *Acanthostega*. We can thus tentatively conclude that *Acanthostega* had a long aquatic juvenile stage characterized, at least at certain times, by the formation of schools that included few or no adults.

Whereas the unique palaeoenvironmental and population-related data provided by the *Acanthostega* mass-death deposit are dependent on the context of that particular locality, the type of life history information provided by the humeri is dependent only on the

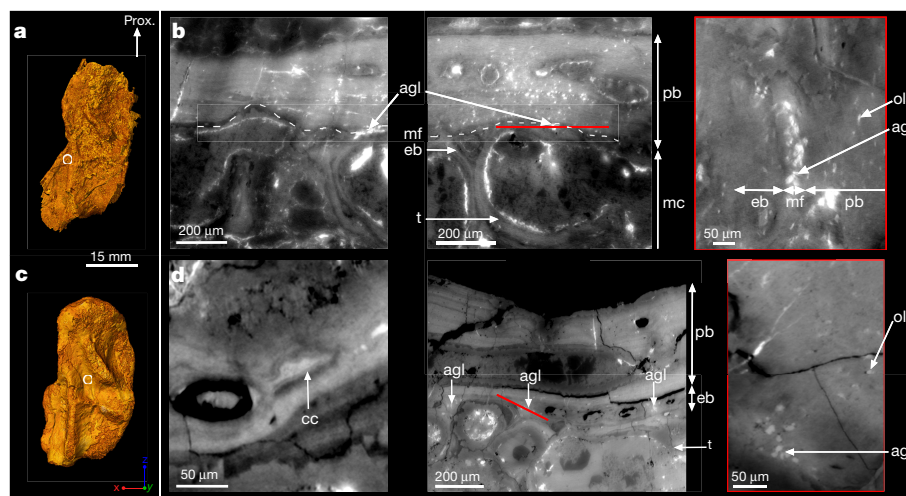


Figure 2 | Humeral bone development. **a**, Three-dimensional model of the humerus of NHMD 74756 in dorsal view. The white circle indicates the midshaft location of the virtual thin sections in **b**. **b**, Humeral cortical histology (voxel size: $0.638\mu\text{m}$; thickness: $10\mu\text{m}$) showing the complete bone deposit of periosteal bone (pb) from the mineralization front (mf) to the surface of the humerus. The aligned globular cell lacunae (agl) are identified as remnants of chondrocyte lacunae, which are much larger than osteocyte lacunae (ol), and typically closely aligned in rows. Trabeculae (t) are numerous in the medullary cavity (mc) and covered with endosteal bone (eb). From left to right: longitudinal virtual thin section; transverse

virtual thin section showing the location of the next section in red; tangential virtual thin section. **c**, Three-dimensional model of the humerus of MGUH 29020 in dorsal view showing the high-resolution scanned location. **d**, Humeral cortical histology (voxel size: $0.638\mu\text{m}$; thickness: $10\mu\text{m}$) showing the complete record of cortical bone deposition. In addition, remnants of calcified cartilage (cc) within the trabeculae explain their endochondral origin. From left to right: longitudinal virtual thin section in the trabecular region; transverse virtual thin section showing the location of the next section in red; oblique virtual thin section.



Figure 3 | Bone skeletochronology. **a**, Three-dimensional models of the humeri of MGUH 29020 (left) and NHMD 74756 (right) in dorsal and ventral views, showing the locations of the high-resolution scans. **b–e**, Virtual thin sections revealing lines of arrested growth (yellow and white arrows) resulting from the cyclical growth of the cortical deposit (c). **b**, Longitudinal section (voxel size: $0.638\mu\text{m}$; thickness: $30\mu\text{m}$). **c**, Transverse section (voxel size: $0.638\mu\text{m}$; thickness: $30\mu\text{m}$). **d**, Transverse section (voxel size: $1.12\mu\text{m}$; thickness: $80\mu\text{m}$). **e**, Transverse section (voxel size: $0.638\mu\text{m}$; thickness: $30\mu\text{m}$). The annual bone growth rate (Extended Data Table 1) was measured in each specimen in regions of cortical bone not distorted by taphonomic or biological factors (such as muscle insertions) and exhibiting regular LAG patterns labelled with white arrows.

preservation of the bone itself and can potentially be matched in a wide range of stem tetrapods. Even a single limb bone can, in principle, provide decisive answers to questions about the length of the juvenile stage and the onset of limb ossification, which in turn help to constrain palaeobiological hypotheses. We are undertaking a systematic PPC-SR μ CT survey of stem tetrapod limb histology with this aim in mind. For now, *Acanthostega* provides a first glimpse of the life history of a Devonian tetrapod.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 March; accepted 10 August 2016.

Published online 7 September 2016.

- Clack, J. A. *Gaining Ground* 2nd edn (Indiana Univ. Press, 2012).
- Niedzwiedzki, G., Szrek, P., Narkiewicz, K., Narkiewicz, M. & Ahlberg, P. E. Tetrapod trackways from the early Middle Devonian period of Poland. *Nature* **463**, 43–48 (2010).
- Shubin, N. H., Daeschler, E. B. & Jenkins, F. A., Jr. The pectoral fin of *Tiktaalik roseae* and the origin of the tetrapod limb. *Nature* **440**, 764–771 (2006).
- Friedman, M., Coates, M. I. & Anderson, P. First discovery of a primitive coelacanth fin fills a major gap in the evolution of lobed fins and limbs. *Evol. Dev.* **9**, 329–337 (2007).
- Boisvert, C. A., Mark-Kurik, E. & Ahlberg, P. E. The pectoral fin of *Panderichthys* and the origin of digits. *Nature* **456**, 636–638 (2008).
- Astin, T. R., Marshall, J. E. A., Blom, H. & Berry, C. M. The sedimentary environment of the Late Devonian East Greenland tetrapods. *Geol. Soc. Lond.* **339**, 93–109 (2010).
- Blom, H., Clack, J. A., Ahlberg, P. E. & Friedman, M. Devonian vertebrates from East Greenland: a review of faunal composition and distribution. *Geodiversitas* **29**, 119–141 (2007).
- Sanchez, S., Ahlberg, P. E., Trinajstić, K. M., Mirone, A. & Tafforeau, P. Three-dimensional synchrotron virtual paleohistology: a new insight into the world of fossil bone microstructures. *Microsc. Microanal.* **18**, 1095–1105 (2012).
- Warburton, F. E. & Denman, N. S. Larval competition and the origin of tetrapods. *Evolution* **15**, 566 (1961).
- Blom, H., Clack, J. A. & Ahlberg, P. E. Localities, distribution and stratigraphical context of the Late Devonian tetrapods of East Greenland. *Medd. Gronl.* **43**, 4–50 (2005).
- Clack, J. A. The dermal skull roof of *Acanthostega gunnari*, an early tetrapod from the Late Devonian. *Trans. R. Soc. Edinb. Earth Sci.* **93**, 17–33 (2002).
- Callier, V., Clack, J. A. & Ahlberg, P. E. Contrasting developmental trajectories in the earliest known tetrapod forelimbs. *Science* **324**, 364–367 (2009).

13. Coates, M. I. The Devonian tetrapod *Acanthostega gunnari* Jarvik: postcranial anatomy, basal tetrapod interrelationships and patterns of skeletal evolution. *Trans. R. Soc. Edinb. Earth Sci.* **87**, 363–421 (1996).
14. Sanchez, S., Tafforeau, P. & Ahlberg, P. E. The humerus of *Eusthenopteron*: a puzzling organization presaging the establishment of tetrapod limb bone marrow. *Proc. R. Soc. Lond. B* **281**, 20140299 (2014).
15. Coates, M. I., Ruta, M. & Friedman, M. Ever since Owen: changing perspectives on the early evolution of tetrapods. *Annu. Rev. Ecol. Syst.* **39**, 571–592 (2008).
16. Francillon-Vieillot, H. *et al.* in *Skeletal Biomineralization: Patterns, Processes and Evolutionary Trends*. (ed. J. G. Carter) Vol. I., 471–530 (Van Nostrand Reinhold, 1990).
17. Laurin, M., Meunier, F.-J., Germain, D. & Lemoine, M. A microanatomical and histological study of the paired fin skeleton of the Devonian sarcopterygian *Eusthenopteron foordi*. *J. Paleontol.* **81**, 143–153 (2007).
18. Castanet, J., Francillon-Vieillot, H. & de Ricqlès, A. in *Amphibian Biology* (eds H. Heatwole & M. Davies) Vol. V Osteology, 1598–1683 (Surrey Beatty & Sons, 2003).
19. Padian, K. Evolutionary physiology: A bone for all seasons. *Nature* **487**, 310–311 (2012).
20. Sanchez, S., Klembara, J., Castanet, J. & Steyer, J.-S. Salamander-like development in a seymouriamorph revealed by palaeohistology. *Biol. Lett.* **4**, 411–414 (2008).
21. Castanet, J., Francillon-Vieillot, H., Meunier, F.-J. & de Ricqlès, A. in *Bone* (ed. B. K. Hall) Vol. 7: Bone Growth B, 245283 (CRC Press, 1993).
22. Fröbisch, N. B. Ossification patterns in the tetrapod limb—conservation and divergence from morphogenetic events. *Biol. Rev. Camb. Philos. Soc.* **83**, 571–600 (2008).
23. Witzmann, F. Developmental patterns and ossification sequence in the Permo-Carboniferous temnospondyl *Archegosaurus decheni* (Saar-Nahe Basin, Germany). *J. Vertebr. Paleontol.* **26**, 717 (2006).
24. Schoch, R. R. Skeleton formation in the Branchiosauridae: a case study in comparing ontogenetic trajectories. *J. Vertebr. Paleontol.* **24**, 309–319 (2004).
25. Peacor, S. D. & Pfister, C. A. Experimental and model analyses of the effects of competition on individual size variation in wood frog (*Rana sylvatica*) tadpoles. *J. Anim. Ecol.* **75**, 990–999 (2006).
26. Badyaev, A. V. Growing apart: an ontogenetic perspective on the evolution of sexual size dimorphism. *Trends Ecol. Evol.* **17**, 369–378 (2002).
27. Kind, P. K. *Movement Patterns and Habitat Use in the Queensland Lungfish Neoceratodus forsteri (Krefft 1870)* PhD Thesis, Univ. Queensland (2002).
28. Sparreboom, M. *Salamanders of the Old World: the Salamanders of Europe, Asia and Northern Africa* (KNNV Publishing, 2014).
29. Ahlberg, P. E. & Milner, A. R. The origin and early diversification of tetrapods. *Nature* **368**, 507–514 (1994).
30. Johanson, Z. The Upper Devonian fish *Bothriolepis* (Placodermi: Antiarchi) from near Canowindra, New South Wales, Australia. *Rec. Aust. Mus.* **50**, 315–348 (1998).

Supplementary Information is available in the online version of the paper.

Acknowledgements Beamtime was allocated as inhouse beamtime and thanks to a proposal accepted by the ESRF (EC203, S.S.). This research was supported by an ERC grant (233111, P.E.A.) and a grant from the Vetenskapsrådet (2015-04335, S.S.). The authors thank J. Castanet, J.-S. Steyer, G. Clement, M. Coates, T. Smithson, A. R. Milner, H. Blom, D. Snitting, I. Adameyko, A. Soler, S. Martin and R. R. Schoch for discussions; G. Cuny and B. E. Kramer Lindow for access to the collections housed in the Natural History Museum of Denmark; and M. Lowe for access to the collections of the University Museum of Zoology, Cambridge.

Author Contributions S.S., P.E.A. and P.T. conceived and designed the project. S.S. and P.T. performed the synchrotron experiments. The localities were excavated by J.A.C. and P.E.A. in 1987. P.T. processed and reconstructed the raw PPC-SRμCT scan data. S.S. segmented the scan data. S.S., P.E.A. and P.T. analysed the data. All authors discussed the interpretations. S.S. and P.E.A. developed the main text. S.S. made the figures and supplementary information. All authors provided a critical review of the manuscript and approved the final draft.

Author Information The synchrotron data will be made available through the ESRF palaeontology database (<http://paleo.esrf.eu>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.S. (sophie.sanchez@ebc.uu.se).

Reviewer Information *Nature* thanks J. Anderson, N. Fröbisch, R. Schoch and K. Stein for their contribution to the peer review of this work.

METHODS

Sampling protocol. All the humeri of *Acanthostega* known so far from the Upper Devonian locality of Stensiö Bjerg (East Greenland) were brought together from museum collections to perform the current study (Natural History Museum of Denmark: MGUH 29019, MGUH 29020, NHMD 74756, University Museum of Zoology Cambridge: UMZC T.1295). They were all imaged at both low and high resolutions at beamline ID19 of the ESRF (see experimental parameters below).

In the diaphyseal region, nine homologous regions were investigated at high resolution in the four humeri but only regions 2, 3, 7, 8 and 9 could provide quantifiable information regarding growth patterns (Extended Data Figs 4, 6 and Extended Data Table 1). Areas of muscle insertion³¹ were avoided as much as possible. Most of the regions (2, 3, 9) are non-muscle attachment areas. Regions 7 and 8 are located between two regions of muscle insertions but their LAG patterns remain undisturbed (Fig. 3e and Extended Data Fig. 5a). Only the LAG pattern of region 7 of MGUH 29020 (Extended Data Fig. 5b) presents an inner cortex that is highly vascularised. In this case, measurements were done in the most external part of the cortex exhibiting a regular untouched LAG pattern.

In the epiphyseal region, one scan was performed at high resolution in specimen MGUH 29020.

Imaging experiments. X-ray imaging was done using propagation phase-contrast synchrotron microtomography (PPC-SRμCT) at beamline ID19 of the ESRF. A multiscale approach allowed us to perform scans with voxel sizes varying from 20.24 to 0.638 μm with average energies ranging from 60 to 123 keV.

Low-resolution experiments. The scan of MGUH 29020, at 20.24 μm voxel size, was done with a monochromatic beam, using a double Si111 Bragg monochromator and a FrelON 2k14 CCD detector³² mounted on the lens coupling optics. The optical system was associated to a Gadox scintillator of 20 μm thickness. The distance between the sample and detector was 950 mm. The scans were performed at 60 keV.

NHMD 74756 was imaged with a voxel size of 14.95 μm with a monochromatic beam, using a double Si111 Bragg monochromator and a 2k14 CCD detector³¹. The optical system was associated to a Gadox scintillator of 10 μm thickness. The sample was positioned 900 mm from the detector. The scans were performed at 60 keV. 5,000 projections were taken over 360° in half-acquisition mode. The time of exposure was of 0.25 s.

In order to obtain a voxel size of 12.62 μm while scanning humeri MGUH 29019, MGUH 29020 and UMZC T.1295, the experiment was done using pink beam with a Frelon 2k14 CCD detector³¹ mounted on an optical system associated to a 1,000-μm-thick LuAG scintillator. The gap of the wiggler was opened to 50 mm. The beam

was filtered with 2 mm of aluminium and 9 mm of copper. The resulting average energy was of 123 keV. The samples were placed 13 m from the detector in order to obtain an enhanced propagation phase-contrast. 4,998 projections were taken over 360° in half-acquisition mode. The time of exposure was of 0.15 s.

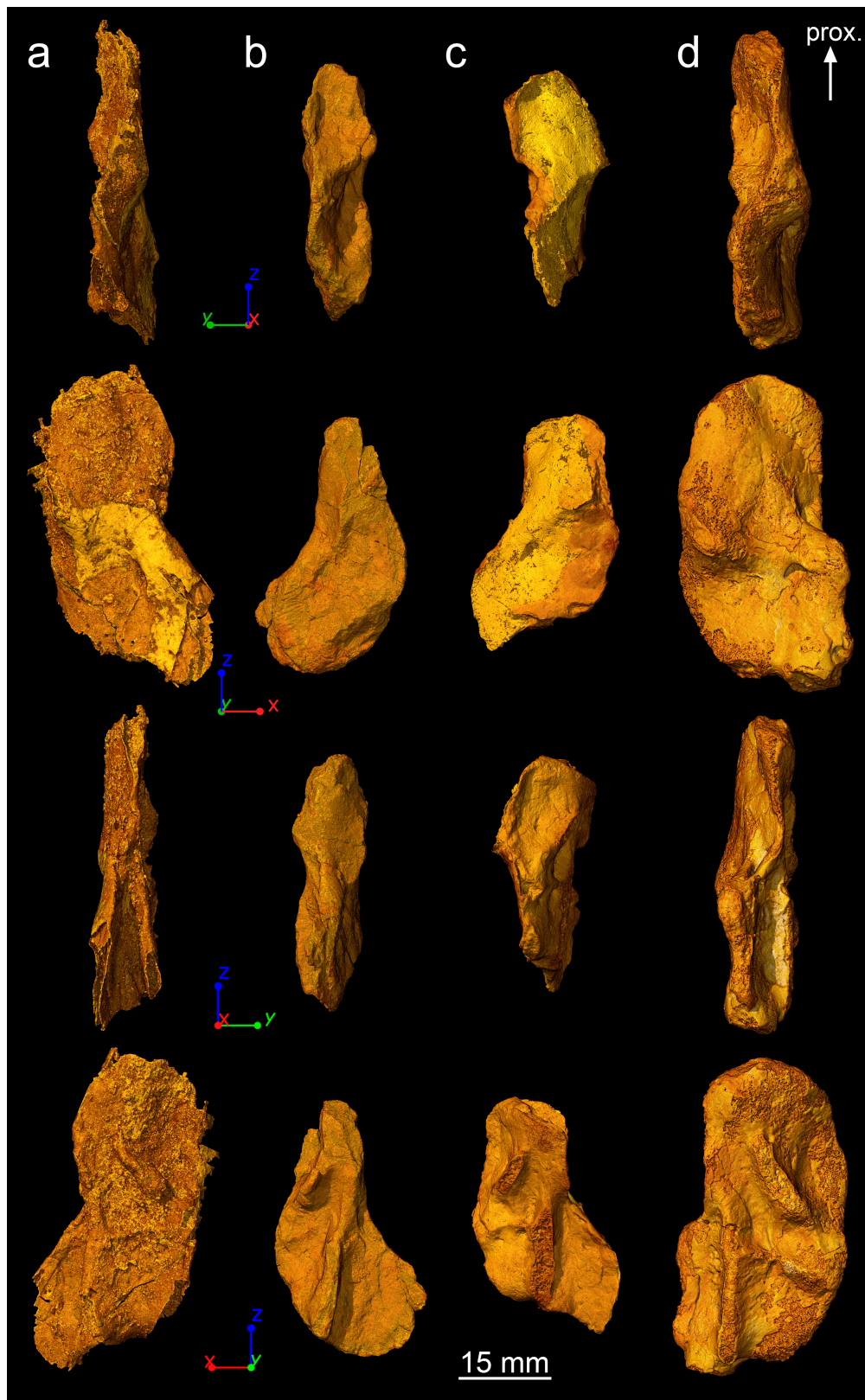
High-resolution experiments. The epiphysis of MGUH 29020 was imaged with a voxel size of 1.12 μm. The experiment was performed in monochromatic conditions with a FrelON 2K14 (ref. 32), a 10× objective, N.A. 0.3, coupled with a 2.5× eyepiece and a 10-μm-thick GGG scintillator mounted on the microscope optics in binning conditions. The multilayer was set to the energy of 52 keV. The distance of propagation was of 150 mm.

The high-resolution scans at midshaft were done with a voxel size of 0.638 μm using pink beam. A Frelon 2k14 CCD detector³² associated to a microscope with a 25 μm-thick GGG scintillator allowed us to obtain a sub-micrometre voxel size. The gap of the U32u undulator was 11.5 mm. The beam was filtered with 2 mm of aluminium, 0.1 mm of copper and 0.1 mm of tungsten. The average energy was probably 55 keV. Twenty-two translocator 2D CRL lenses, made of beryllium and with a radius of curvature of 0.5 mm, were used to reduce the divergence of the beam at this energy, thereby increasing the flux going through the samples. The fossils were at a distance of propagation of 150 mm. 6,000 projections were taken over 360° in half-acquisition mode. The time of exposure was 1 s.

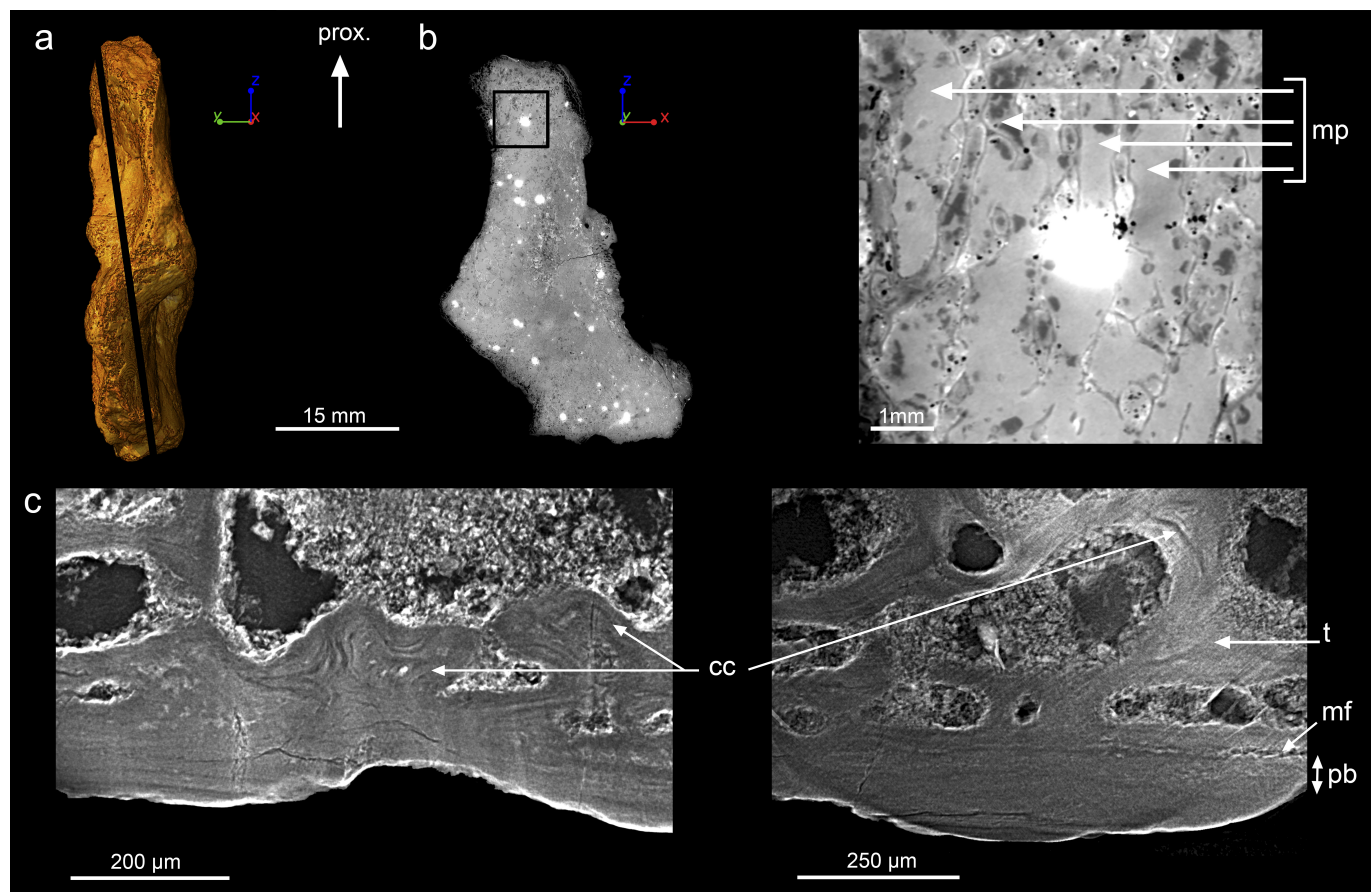
Image reconstruction. The data were reconstructed using a single distance phase retrieval approach⁸ based on a modified version of the algorithm of Paganin *et al.*³³, applying an unsharp mask to the radiographs after the phase retrieval to compensate for the partial loss of high frequencies due to the original algorithm. This is based on a relative chemical homogeneity assumption. In-house filters were used to enhance the contrast between the microstructures and to reduce the noise induced by metallic infillings.

Segmentation. The data were segmented using VGStudio MAX version 2.2 software (Volume Graphics Inc.).

31. Bishop, P. J. The humerus of *Ossinodus pueri*, a stem tetrapod from the Carboniferous of Gondwana, and the early evolution of the tetrapod forelimb. *Alcheringa Australas. J. Palaeontol.* **38**, 209–238 (2014).
32. Labiche, J.-C. *et al.* Invited article: the fast readout low noise camera as a versatile x-ray detector for time resolved dispersive extended x-ray absorption fine structure and diffraction studies of dynamic problems in materials science, chemistry, and catalysis. *Rev. Sci. Instrum.* **78**, 091301–091311 (2007).
33. Paganin, D., Mayo, S. C., Gureyev, T. E., Miller, P. R. & Wilkins, S. W. Simultaneous phase and amplitude extraction from a single defocused image of a homogeneous object. *J. Microsc.* **206**, 33–40 (2002).

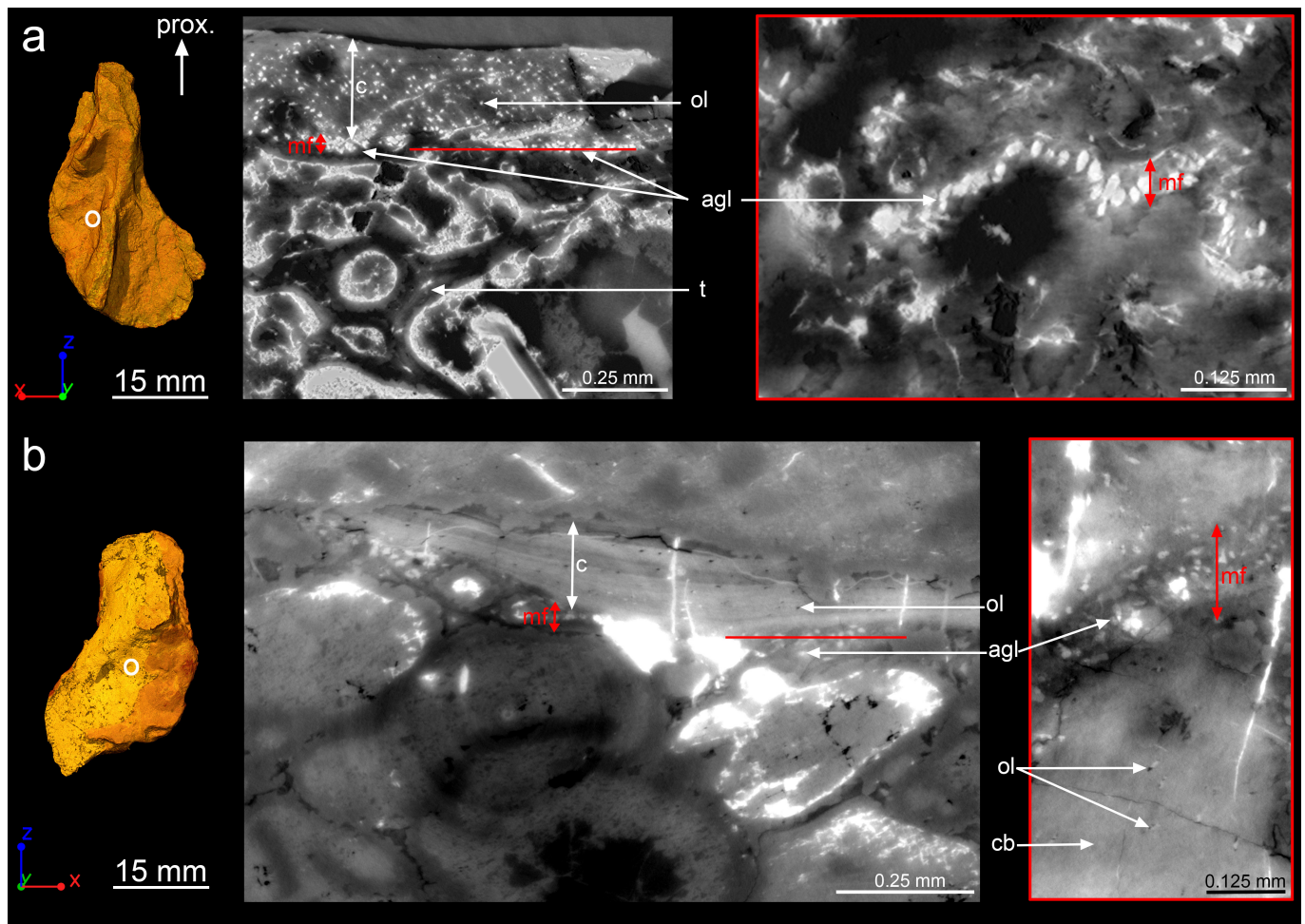


Extended Data Figure 1 | Three-dimensional models of *Acanthostega* humeri based on synchrotron microtomography data. a, NHMD 74756. b, UMZC T.1295. c, MGUH 29019. d, MGUH 29020. From top to bottom: preaxial view, ventral view, postaxial view, dorsal view. Humeri are all oriented with their proximal epiphysis towards the top.



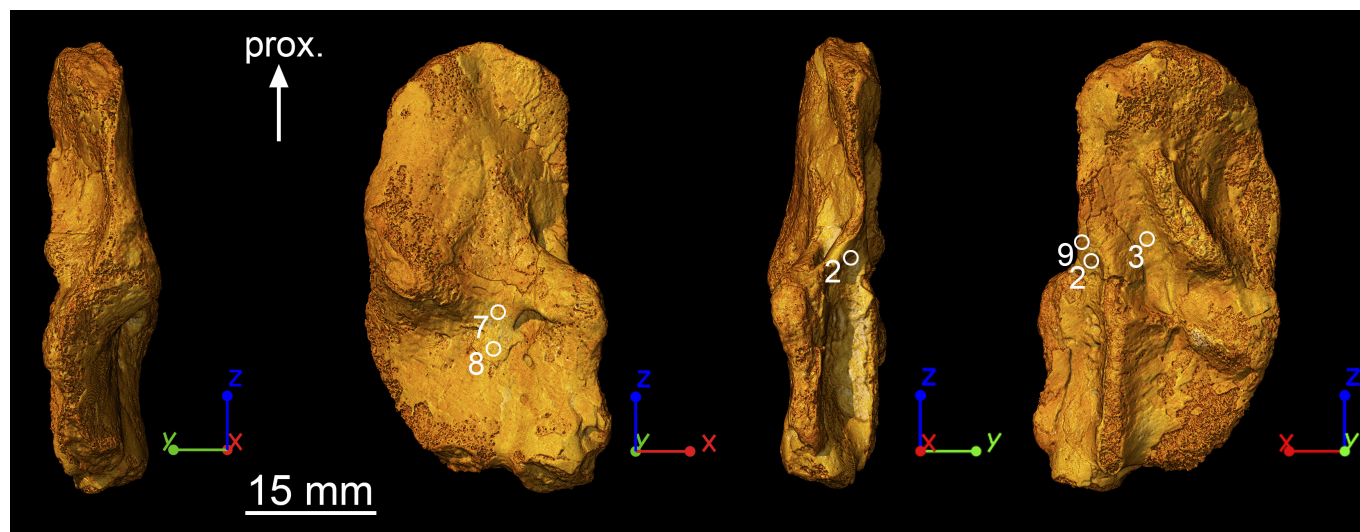
Extended Data Figure 2 | Epiphyseal microanatomy and histology of *Acanthostega* humerus (MGUH 29020). **a**, Three-dimensional model in preaxial view, based on synchrotron microtomography data, oriented with the proximal extremity (epiphysis⁶) towards the top. The black line indicates the virtual thin section illustrated in **b**. **b**, Longitudinal virtual thin section (thickness: 50 μm , voxel size: 1.12 μm , same scale bar and orientation as in **a**) showing the location of the detailed image on the right. The latter shows the marrow processes (mp) formed in the growth

plate by endochondral ossification. **c**, High-resolution virtual thin section (thickness: 50 μm , voxel size: 1.12 μm) from the epiphyseal region showing obvious Liesegang's rings as remnants of calcified cartilage⁶ (cc), formed during endochondral ossification. These remnants are entrapped in the trabeculae (t), at the vicinity of the ossification notch⁶, where the thickness of the periosteal bone (pb) between the mineralization front (mf) and the surface is greatly reduced. The bone is oriented with its surface towards the bottom. Left, longitudinal thin section; right, transverse section.



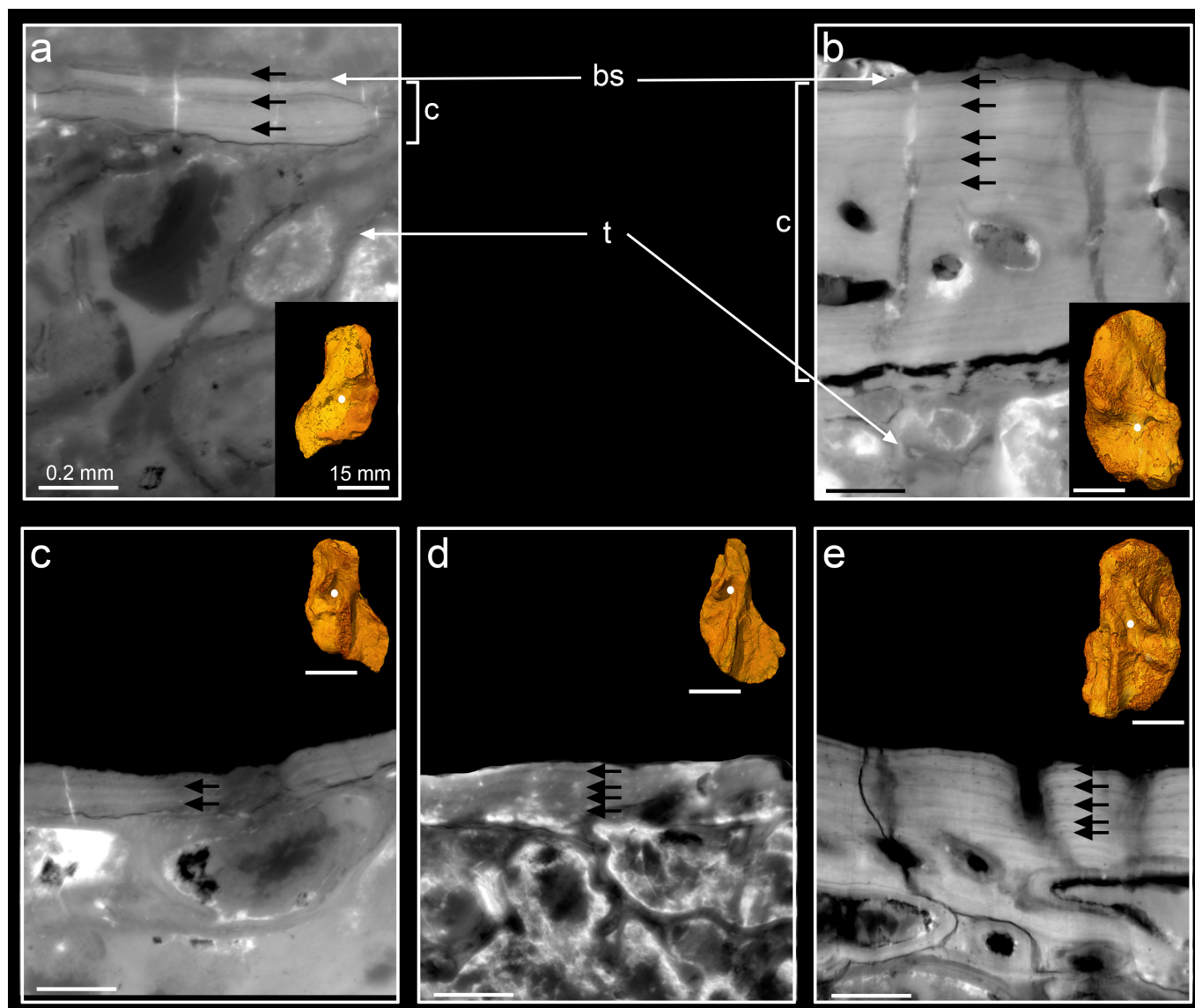
Extended Data Figure 3 | Midshaft bone histology of two *Acanthostega* humeri (UMZC T.1295 and MGUH 29019). **a**, Three-dimensional model of humerus UMZC T.1295 in dorsal view and oriented with the proximal epiphyses⁶ towards the top. The white circle indicates the midshaft location at which the transverse virtual section was made. The latter (single tomographic slice, voxel size: $0.638\mu\text{m}$) shows the complete bone deposit of cortical bone (c) from the mineralization front (mf) to the surface of the humerus (top). The cortical bone comprises numerous osteocyte lacunae (ol), which are much smaller than the aligned globular cell lacunae (agl) present at the location of the mineralization front. Trabeculae (t) are numerous in the medullary cavity. The red

line in the transverse virtual section indicates the location of the next tangential virtual section which details the mineralization front. **b**, Three-dimensional model of the humerus MGUH 29019 in ventral view showing the high-resolution scanned location. The virtual section shows the humeral cortical histology at the midshaft (single tomographic slice, voxel size: $0.638\mu\text{m}$). As in UMZC T.1295, the cortical bone matrix (cb) is very compact, pierced with small osteocyte lacunae. At this location, its surface (top), although still embedded in the rock matrix, is not well preserved. The red line in the transverse virtual section indicates the location of the next tangential virtual section detailing the cellular structure of the mineralization front.



Extended Data Figure 4 | Regions of high-resolution scans. Skeletochronological observations were done at sub-micrometre resolution in nine homologous regions of the four humeri of *Acanthostega*. Specimen MGUH 29020 is used here to illustrate the regions providing quantifiable information to calculate annual bone growth rates (Extended

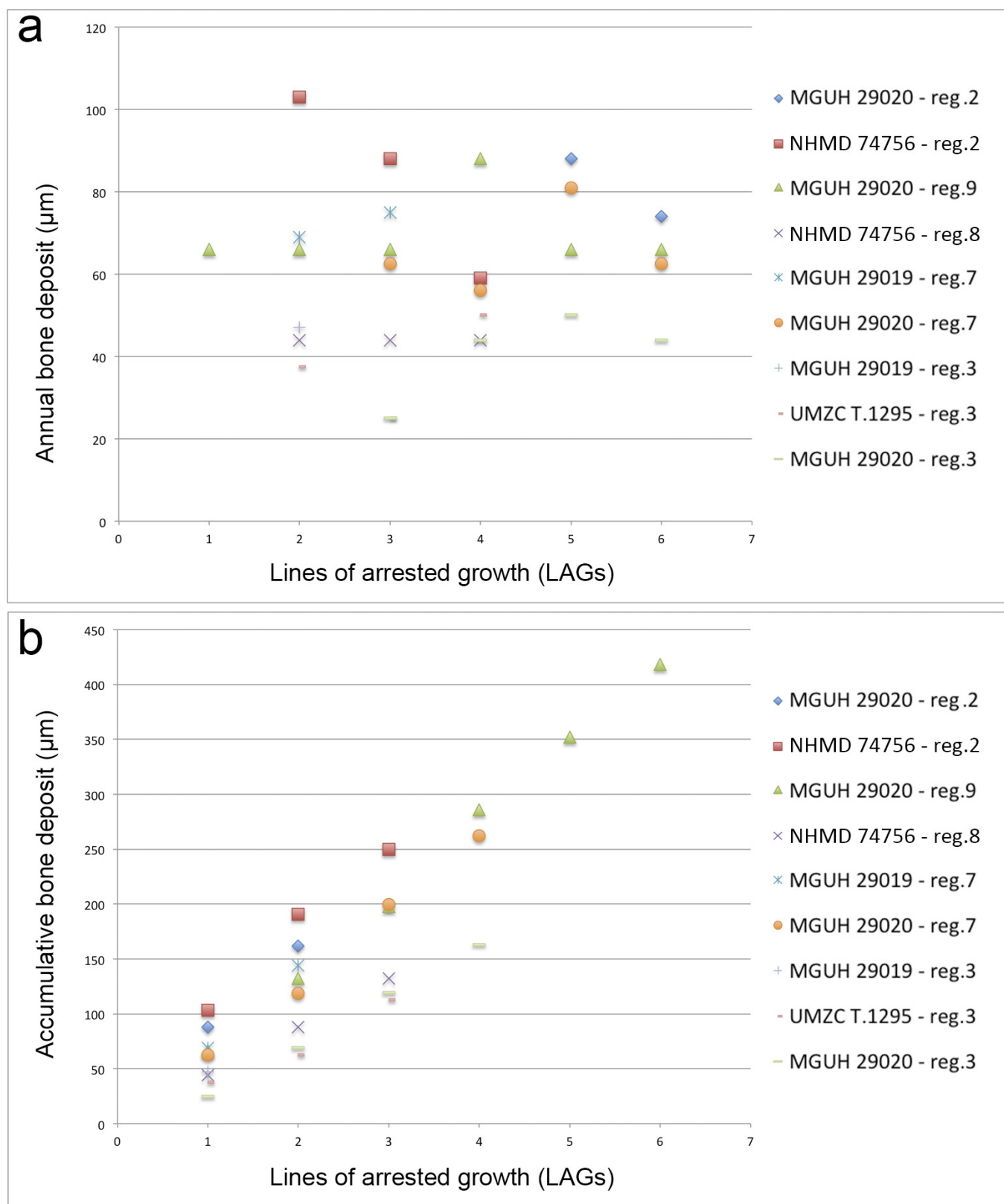
Data Table 1). Areas of muscle insertion were avoided when possible. Regions 2, 3 and 9 are non-muscle attachment areas. Regions 7 and 8 are located between two regions of muscle insertions but annual bone growth rates (Extended Data Table 1) were measured only in undisturbed cortical parts exhibiting regular LAG patterns.



Extended Data Figure 5 | Humeral midshaft skeletochronology.

All virtual thin sections (voxel size: $0.638\mu\text{m}$) reveal LAGs (black arrows) resulting from the cyclical growth of the cortical deposit (c). They are oriented with the surface of the bone (sb) towards the top and medullary trabeculae (t) downwards. The locations of the thin sections are shown as white dots on the associated 3D models. All 3D models are oriented with their proximal epiphyses^o towards the top. **a**, Transverse virtual thin section (thickness: $30\mu\text{m}$) showing three LAGs in the cortical bone of the ventral midshaft of the humerus MGUH 29019 (region 7). The inner surface of the cortical bone has been eroded. **b**, Longitudinal virtual thin section (thickness: $30\mu\text{m}$) showing five LAGs in the cortical bone of the ventral midshaft of MGUH 29020 (region 7). The inner cortical bone is disturbed by a highly vascularised period. LAGs cannot be identified with

accuracy in this region. The growth deposits between the LAGs in region 7 are similar in MGUH 29019 and MGUH 29020 (Extended Data Table 1). **c**, Transverse virtual thin section (thickness: $30\mu\text{m}$) showing two LAGs in the cortical bone of the dorsal midshaft of the specimen MGUH 29019 (region 3). **d**, Longitudinal virtual thin section (thickness: $50\mu\text{m}$) showing four LAGs in the cortical bone of the dorsal midshaft of UMZC T.1295 (region 3). **e**, Longitudinal virtual thin section (thickness: $30\mu\text{m}$) showing five LAGs in the cortical bone of the dorsal midshaft of MGUH 29020 (region 3). The growth deposits between the LAGs in region 3 are similar in UMZC T.1295, MGUH 29019 and MGUH 29020 (Extended Data Table 1). Scale bars for virtual thin sections: 0.2 mm . Scale bars for 3D models: 15 mm .



Extended Data Figure 6 | Graphic visualizations of bone deposits. Images are based on the measurements provided in Extended Data Table 1. **a**, Amount of bone deposited every year—that is, between two LAGs—in the regions of interest (reg.) of the four studied humeri. Except for region 2 (measured in MGUH 29020 and NHMD 74756), all regions show a

relatively constant or increasing growth rate during animal development. **b**, Bone deposition accumulated to form the cortex. Despite a slight variation in values due to growth allometries, the growth rate (illustrated by the slope angle) is relatively constant in all regions of all specimens, meaning that all specimens grew at the same rate.

Extended Data Table 1 | Measurements of humeral cyclical growth deposits between LAGs

Region	Specimen	Figure	LAG 0-1	LAG 1-2	LAG 2-3	LAG 3-4	LAG 4-5	LAG 5-6	Average deposit
2	MGUH 29020	Fig. 3b	LAG pattern disturbed in a highly vascularised region				88	74	81
2	NHMD 74756	Fig. 3c	eroded	103	88	59	-	-	83
9	MGUH 29020	Fig. 3d	66	66	66	88	66	66	70
8	NHMD 74756	Fig. 3e	eroded	44	44	44	-	-	44
7	MGUH 29019	Extended Data Fig. 5a	eroded	69	75	-	-	-	72
7	MGUH 29020	Extended Data Fig. 5b	LAG pattern disturbed in a highly vascularised region		62.5	56	81	62.5	65.5
3	MGUH 29019	Extended Data Fig. 5c	eroded	47	-	-	-	-	47
3	UMZC T.1295	Extended Data Fig. 5d	eroded	37.5	25	50	-	-	37.5
3	MGUH 29020	Extended Data Fig. 5e	LAG pattern disturbed in a highly vascularised region		25	44	50	44	41

Measurements (in μm) are based on the LAGs labelled with white arrows in Fig. 3 and black arrows in Extended Data Fig. 5. Nine regions were investigated but only regions 2, 3, 7, 8 and 9 could provide quantifiable information (Extended Data Fig. 4). A dash indicates that there were no more LAGs as the bone had stopped growing.

Follicular CXCR5-expressing CD8⁺ T cells curtail chronic viral infection

Ran He^{1*}, Shiyue Hou^{2*}, Cheng Liu¹, Anli Zhang³, Qiang Bai¹, Miao Han⁴, Yu Yang³, Gang Wei⁴, Ting Shen⁴, Xinxin Yang¹, Lifan Xu¹, Xiangyu Chen¹, Yaxing Hao¹, Pengcheng Wang¹, Chuhong Zhu⁵, Juanjuan Ou⁶, Houjie Liang⁶, Ting Ni⁴, Xiaoyan Zhang³, Xinyuan Zhou¹, Kai Deng⁷, Yaokai Chen⁸, Yadong Luo⁸, Jianqing Xu³, Hai Qi², Yuzhang Wu¹ & Lilin Ye¹

During chronic viral infection, virus-specific CD8⁺ T cells become exhausted, exhibit poor effector function and lose memory potential^{1–4}. However, exhausted CD8⁺ T cells can still contain viral replication in chronic infections^{5–9}, although the mechanism of this containment is largely unknown. Here we show that a subset of exhausted CD8⁺ T cells expressing the chemokine receptor CXCR5 has a critical role in the control of viral replication in mice that were chronically infected with lymphocytic choriomeningitis virus (LCMV). These CXCR5⁺ CD8⁺ T cells were able to migrate into B-cell follicles, expressed lower levels of inhibitory receptors and exhibited more potent cytotoxicity than the CXCR5[−] subset. Furthermore, we identified the Id2–E2A signalling axis as an important regulator of the generation of this subset. In patients with HIV, we also identified a virus-specific CXCR5⁺ CD8⁺ T-cell subset, and its number was inversely correlated with viral load. The CXCR5⁺ subset showed greater therapeutic potential than the CXCR5[−] subset when adoptively transferred to chronically infected mice, and exhibited synergistic reduction of viral load when combined with anti-PD-L1 treatment. This study defines a unique subset of exhausted CD8⁺ T cells that has a pivotal role in the control of viral replication during chronic viral infection.

During chronic viral infection, exhausted CD8⁺ T cells remain able to mediate imperative control of viral replication in both animal models and human immunodeficiency virus (HIV) infection^{5–9}. Here we investigate whether the exhausted CD8⁺ T-cell pool constitutes a specific subset that effectively controls viral replication, and whether a specific niche within secondary lymphoid tissues accommodates this subset. To this end, we infected mice with either the lymphocytic choriomeningitis (LCMV) Armstrong strain (resulting in acute infection) or the LCMV clone 13 (Cl13) strain (resulting in chronic infection). Notably, we visualized a substantial accumulation of CD8⁺ T cells in B-cell follicles in chronically, but not acutely, infected mice (Fig. 1a, b). As CXCR5 directs B and T lymphocytes to localize to B-cell follicles^{10–12}, we found that approximately 30% of LCMV-specific CD8⁺ T cells highly expressed CXCR5 in lymphoid tissues from chronically, but not acutely, infected mice (Fig. 1c, Extended Data Fig. 1a). Furthermore, the population of virus-specific CXCR5⁺ CD8⁺ T cells remained constant in the spleen during chronic infection (Fig. 1d). CXCR5⁺ CD8⁺ T cells were not detected in non-lymphoid tissues after chronic infection (Extended Data Fig. 1b). Transferred CXCR5⁺ CD8⁺ T cells migrated into the B-cell follicles in infection-matched CD8-knockout (*Cd8*^{−/−}) recipients, while their CXCR5[−] counterparts mostly failed to do so (Fig. 1e). CXCR5⁺ CD8⁺ T cells probably do not represent the non-classical MHC Q-1a-restricted regulatory

CD8⁺ T-cell population, as they did not express ICOS ligand and transcription factor Helios^{13,14} (Extended Data Fig. 1c). Together, these data demonstrate a subset of virus-specific CD8⁺ T cells expressed CXCR5 and migrated to B-cell follicles during chronic viral infection.

Next, we compared the phenotype and function of virus-specific CXCR5⁺ and CXCR5[−] CD8⁺ T cells. Notably, CXCR5⁺ CD8⁺ T cells expressed lower levels of inhibitory molecules PD-1 and Tim-3 and higher levels of the stimulatory molecule KLRG1 than their CXCR5[−] counterparts (Fig. 2a, b, Extended Data Fig. 2a, b). Moreover, the CXCR5⁺ subset produced higher levels of interferon γ (IFN- γ) and tumour necrosis factor α (TNF α) than the CXCR5[−] subset upon re-stimulation, although a comparable frequency of cytokine-producing cells was shown in both subsets at the early stage of exhaustion (day 8 after infection) (Fig. 2c, Extended Data Fig. 2c). Notably, approximately 30% of CXCR5⁺ CD8⁺ T cells were positive for the surface-staining of the degranulation marker CD107, whereas the CXCR5[−] subset was almost negative for this marker (Fig. 2d), and consistently, these cells were more efficient at killing target cells than the CXCR5[−] subset (Fig. 2e). To test the antiviral effect of the CXCR5⁺ subset *in vivo*, we transferred sorted CD44^{hi} CXCR5⁺ and CD44^{hi} CXCR5[−] CD8⁺ T cells into chronically infected *Cd8*^{−/−} recipients. We found that viral load in lungs and spleens of recipients that received CXCR5⁺ donor cells was almost 1,000-fold lower than in tissues of recipients engrafted with CXCR5[−] donor cells (Fig. 2f). These data together support the notion that the virus-specific CXCR5⁺ CD8⁺ T-cell subset is less exhausted and possesses higher antiviral potential *in vivo* than the CXCR5[−] subset.

We further examined whether CXCR5⁺ CD8⁺ T cells affected germinal centre B cells and follicular helper T (T_{FH}) cells. To this end, sorted CXCR5⁺ and CXCR5[−] CD8⁺ T cells were transferred into infection-matched *Cd8*^{−/−} recipients, and we observed comparable germinal centre B- and T_{FH}-cell populations and similar LCMV-specific IgG levels between recipients that received either subset (Extended Data Fig. 3a–c). These results indicated that CXCR5⁺ CD8⁺ T cells had a minimal effect on germinal centre B- and T_{FH}-cell responses. Notably, we found that both naive and germinal centre B cells exhibited lower expression levels of inhibitory molecules programmed death-ligand 1 and 2 (PD-L1 and PD-L2) than other cell types that primarily reside in the T-cell zone (Extended Data Fig. 3d). To elucidate the role of the follicle structure in supporting the CXCR5⁺ CD8⁺ T-cell population, we transferred CXCR5⁺ CD8⁺ T cells to infection-matched wild-type mice or B-cell-deficient mice (hereafter, μ MT mice) and found that the transferred cells were poorly maintained in μ MT recipients compared to wild-type recipients (Extended Data Fig. 4a–c). Moreover,

¹Institute of Immunology, Third Military Medical University, Chongqing 400038, China. ²Tsinghua-Peking Center for Life Sciences, Laboratory of Dynamic Immunobiology, School of Medicine, Tsinghua University, Beijing 100084, China. ³Shanghai Public Health Clinical Center & Institutes of Biomedical Sciences, Fudan University, Shanghai 201508, China. ⁴State Key Laboratory of Genetic Engineering & MOE Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center of Genetics and Development, School of Life Sciences, Fudan University, Shanghai 200438, China. ⁵Department of Anatomy, School of Basic Medicine, Third Military Medical University, Chongqing 400038, China. ⁶Department of Oncology, Southwestern Hospital, Third Military Medical University, Chongqing 400038, China. ⁷Institute of Human Virology, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510080, China. ⁸Chongqing Public Health Medical Center, Chongqing 400000, China.

*These authors contributed equally to this work.

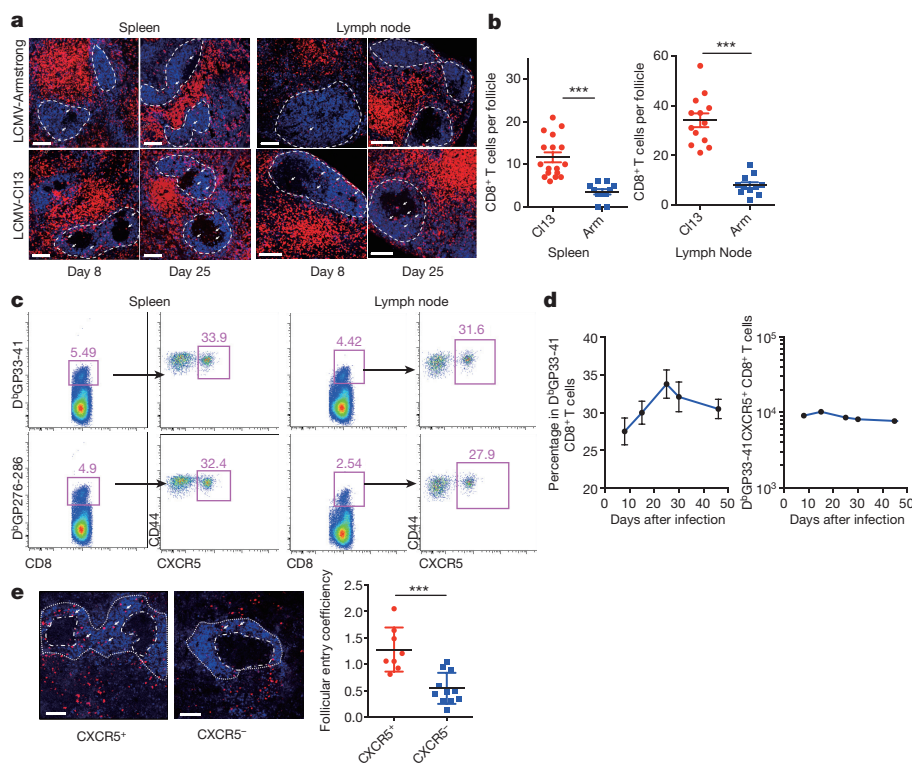


Figure 1 | Virus-specific CXCR5⁺CD8⁺ T cells are generated during chronic infection and migrate into B-cell follicles. **a**, CD8⁺ T-cell localization in the spleens and lymph nodes of mice infected with LCMV Armstrong (Arm⁺) and LCMV Cl13 (Cl13) on days 8 and 25 after infection (blue, IgD; red, CD8). **b**, The number of CD8⁺ T cells per follicle viewed on day 25 after Cl13 infection (spleen, Cl13, $n = 17$, Arm⁺, $n = 11$; lymph node, Cl13, $n = 13$, Arm⁺, $n = 11$). **c**, CXCR5 expression in tetramer-specific CD8⁺ T cells as viewed on day 8 after Cl13 infection. **d**, Kinetics of tetramer-specific CXCR5⁺CD8⁺ T cells during Cl13

infection in the spleen ($n = 4$). **e**, Equal numbers of sorted CXCR5⁺ and CXCR5⁻CD8⁺ T cells were adoptively transferred into infection-matched day-8-infected CD8^{-/-} mice, followed by confocal microscopy analysis with spleen sections on day 5 after transfer (blue, IgD; red, CD8) and follicular entry coefficient was calculated (CXCR5⁺, $n = 8$; CXCR5⁻, $n = 11$). Scale bar (**a**, **e**), 100 μ m. The data are representative of two (**a**, **e**) or three (**c**, **d**) independent experiments, and were analysed by two-tailed unpaired *t*-test (**b**, **e**). Error bars (**b**, **d**, **e**) denote s.e.m. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

the cytolytic capacity and cytokine production of the transferred cells was also compromised in μ MT recipients (Extended Data Fig. 4d, e). Consistently, transferred CXCR5⁺CD8⁺ T cells inhibited viral replication more efficiently in wild-type recipients than in μ MT recipients (Extended Data Fig. 4f). These results together suggest that B-cell follicles represent a specialized niche that accommodates the less-exhausted CXCR5⁺CD8⁺ T cells and preserves their effector functions during chronic viral infection.

To investigate whether CXCR5 expression by virus-activated CD8⁺ T cells is required for their follicular localization, we constructed splenic chimaeric mice in which CD8⁺ T cells were deficient for CXCR5, whereas in most of the other immune cells, CXCR5 expression was intact (Extended Data Fig. 5a). On day 15 after infection, CXCR5-deficient CD8⁺ T cells could barely be detected in the B-cell follicles and exhibited lower CD107 expression and IFN- γ secretion than wild-type counterparts (Extended Data Fig. 5b, c). Consistently, the viral titers were lower in recipients reconstituted with splenocytes derived from wild-type and $Cd8^{-/-}$ mice than those reconstituted with $Cxcr5^{-/-}$ and $Cd8^{-/-}$ splenocytes (Extended Data Fig. 5d). These data demonstrate the importance of cell-autonomous CXCR5 expression in virus-activated CD8⁺ T cells for their follicular localization.

To elucidate the differences in cellular process and functional states between the two subsets, we performed RNA sequencing (RNA-seq). The gene expression pattern in the CXCR5⁺ subset differed greatly from that in the CXCR5⁻ subset at a genome-wide level. Specifically, genes encoding TNF family proteins and their receptors and certain chemokine receptors were more enriched in the CXCR5⁺ subset than in the CXCR5⁻ subset (Extended Data Fig. 6a), and these results were further corroborated by Gene Ontology analysis (Extended Data Fig. 6b).

These results suggest that CXCR5⁺ and CXCR5⁻ cells may represent two subsets of cytotoxic T cells with distinct cell states. We further identified 13 transcription-binding motifs potentially associated with the differentially expressed genes between these two subsets (Extended Data Fig. 6c). Notably, V\$E47_02 (a binding motif associated with the transcription factor E2A isoform, E47) was significantly enriched in the CXCR5⁺ subset (Extended Data Fig. 6c). Consistently, CXCR5⁺ cells expressed the lower levels of Id2, which antagonizes E2A transcriptional activity¹⁵, than CXCR5⁻ cells, albeit with a comparable E2A expression level between two subsets (Extended Data Fig. 6d, e, 7a).

Next, we investigated whether the Id2–E2A signalling axis regulated CXCR5⁺CD8⁺ T-cell differentiation during chronic infection. We infected $Cd4^{Cre}$ - $Id2^{fl/fl}$ (termed $Id2^{-/-}$) mice, whose CD8⁺ T cells expressed less than 1% of Id2 compared to cells from littermate controls (Extended Data Fig. 7b). On day 21 after infection, both mice displayed a comparable expansion of CD44^{hi}CD8⁺ T cells (Extended Data Fig. 7c), whereas the frequency and number of CXCR5⁺CD8⁺ T cells were significantly higher in $Id2^{-/-}$ mice than in controls (Fig. 3a). Moreover, Id2-null CXCR5⁺CD8⁺ T cells exhibited reduced inhibitory molecule expression and more potent effector function than control cells (Fig. 3b). Accordingly, the virus titers were also significantly lower in $Id2^{-/-}$ mice than in control mice (Fig. 3c). These phenotypes were recapitulated in mixed bone marrow chimaeras, reconstituted with donor cells derived from $Id2^{-/-}$ and wild-type mice (Fig. 3d, e). E2A has been reported to regulate CXCR5 expression in CD4⁺ T cells^{16,17}. We further determined whether *Cxcr5* expression was directly targeted by the E2A–Id2 axis in CXCR5⁺CD8⁺ T cells generated during chronic viral infection. By screening the published chromatin immunoprecipitation (ChIP)-sequencing data¹⁶, we identified a conserved consensus

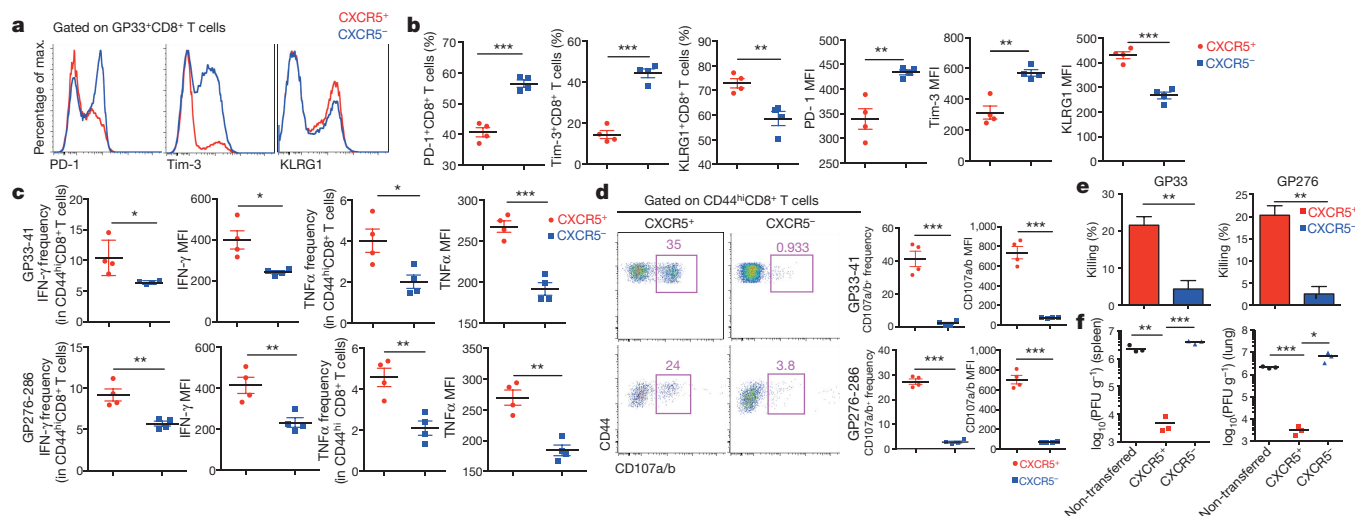


Figure 2 | Virus-specific CXCR5⁺ CD8⁺ T cells are less exhausted than CXCR5⁻ CD8⁺ T cells and control viral load during chronic infection. **a, b**, PD-1, Tim-3 and KLRG1 expression in virus-specific CXCR5⁺ and CXCR5⁻ CD8⁺ T cells from the spleens of Cl13-infected mice on day 25 after infection ($n = 4$). MFI, mean fluorescence intensity. **c, d**, Upon stimulation with the indicated peptides, the cytokine production and CD107 surface expression of virus-specific CXCR5⁺ and CXCR5⁻ CD8⁺ T cells from the spleens of Cl13-infected mice were examined on day 25 after infection ($n = 4$). **e**, *In vivo* killing efficiency of virus-specific

CXCR5⁺ and CXCR5⁻ CD8⁺ T cells from day 21 Cl13-infected mice ($n = 3$). **f**, Equal numbers of CD44^{hi}CXCR5⁺ and CD44^{hi}CXCR5⁻ CD8⁺ T cells sorted from day 8 Cl13-infected mice were adoptively transferred into infected *Cd8^{-/-}* recipients at two intervals (days 21 and 28 after recipient infection). Five days after the final transfer, virus titration of the indicated tissues was performed ($n = 3$). The data are representative of three (**a–e**) or two (**f**) independent experiments, and were analysed by two-tailed unpaired *t*-test (**b–f**). Error bars (**b–e**) denote s.e.m. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

E2A-binding sequence in the *Cxcr5* intron region (Extended Data Fig. 7d). This putative binding site was confirmed by ChIP with quantitative PCR (ChIP-qPCR) with sorted CXCR5⁺ CD8⁺ T cells (Fig. 3f) and was potentially permissive for E2A binding owing to the preferential histone modification with me3H3K4, but not me3H3K27 (Fig. 3g). Using a self-inactivating retroviral reporter system, we noted that the vector containing the mutant E2A-binding motif transcribed much less of the downstream reporter (Thy-1.1) than the vector containing wild-type motif (Extended Data Fig. 7e, f). These data together indicated that E2A promoted *Cxcr5* expression via binding to its intron region in exhausted CD8⁺ T cells. Consistently, over-expressing E2A in LCMV-specific P14 cells markedly upregulated CXCR5 expression and the frequency of CXCR5⁺ cells in P14 cells, whilst co-overexpressing Id2 compromised such effect (Extended Data Fig. 7g). Importantly, E2A-overexpressing cells exhibited a diminished PD-1 expression but enhanced surface CD107 expression and cytokine secretion compared to control cells (Extended Data Fig. 7h). These results together demonstrated that the E2A–Id2 axis functions to regulate the differentiation and effector function of CXCR5⁺ CD8⁺ T cells during chronic infection.

Next, we sought to probe whether activated CXCR5⁺ CD8⁺ T cells would convert into CXCR5⁻ CD8⁺ T cells during chronic infection. To this end, we generated CXCR5–GFP knock-in mice, with a GFP reporter for CXCR5 expression (Extended Data Fig. 8a). We sorted GFP⁺CD44^{hi}CD8⁺ T cells and GFP⁻CD44^{hi}CD8⁺ T cells from day-8 Cl13-infected knock-in mice and labelled these cells with cell-division dye and transferred them into infection-matched congenic recipients (Extended Data Fig. 8b). On day 5 after transfer, most of the transferred GFP⁺CD8⁺ T cells had differentiated into GFP⁻CD8⁺ T cells with vigorous division, whereas the transferred GFP⁻CD8⁺ T cells experienced a low level of division and failed to differentiate into GFP⁺ cells (Extended Data Fig. 8c). On day 12 after transfer, all transferred GFP⁺CD8⁺ T cells had ultimately converted into GFP⁻ cells, accompanied by the upregulation of Id2 expression (Extended Data Fig. 8c, d). Notably, the expression of CD107, the production of IFN- γ and cytolytic capacity was hierarchically reduced in GFP⁺CD8⁺ T cells (GFP⁺), GFP⁻CD8⁺ T cells newly derived from GFP⁺CD8⁺ T cells (GFP⁺/GFP⁻) and GFP⁻CD8⁺ T cells (GFP⁻) (Extended Data

Fig. 8e, f). These results indicate that CXCR5⁺ CD8⁺ T cells may serve as progenitors that further differentiate into CXCR5⁻ CD8⁺ progeny cells, with these newly converted cells possessing better effector functions than cells that were always negative for CXCR5 expression during chronic infection.

The conversion of CXCR5⁺ CD8⁺ T cells into CXCR5⁻ CD8⁺ T-cell population was also present in chronic HIV infection. HIV-specific CXCR5⁺ CD8⁺ T cells could indeed be found in the blood of infected patients (Extended Data Fig. 9a). These CXCR5⁺ cells exhibited lower levels of PD-1 and Tim3 expression than CXCR5⁻ cells (Extended Data Fig. 9b). Notably, serum viral load was inversely correlated with blood CXCR5⁺ CD8⁺ T-cell number in patients prior to anti-retroviral therapy (Extended Data Fig. 9c). HIV-specific CXCR5⁺ CD8⁺ T cells were also present in lymph nodes (Extended Data Fig. 9d), which were potentially localized to the B-cell follicles (Extended Data Fig. 9e), consistent with previous reports^{19–21}. These CXCR5⁺ cells secreted higher levels of IFN- γ and TNF α , and expressed higher levels of CD107 and perforin than their CXCR5⁻ counterparts (Extended Data Fig. 9f). Notably, CXCR5⁺ cells exhibited lower levels of Id2 than CXCR5⁻ CD8⁺ T cells, although both populations expressed comparable levels of E2A (Extended Data Fig. 9g). Taken together, these results highlight that both LCMV- and HIV-specific CXCR5⁺ CD8⁺ T cells are able to migrate to B-cell follicles, display higher effector function and, presumably, are similarly programmed to control viral replication more efficiently than CXCR5⁻ CD8⁺ T cells.

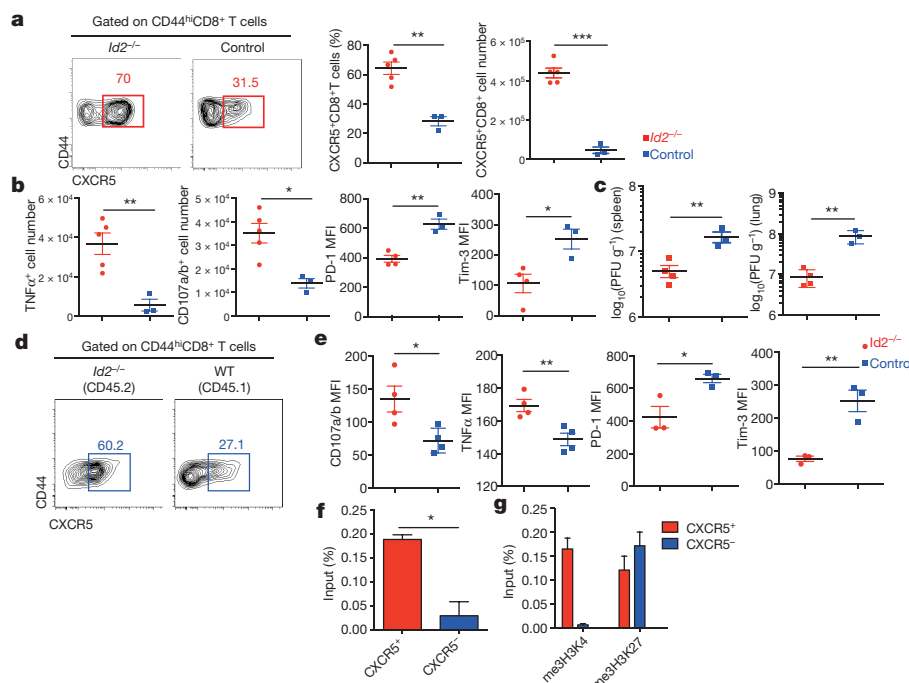


Figure 3 | The Id2–E2A axis is critical for the differentiation of the CXCR5⁺CD8⁺ T-cell subset during chronic viral infection. **a**, Frequency and total number of CXCR5⁺CD8⁺ T cells in the spleens of *Id2*^{−/−} and litter-mate control (control) mice on day 21 after Cl13 infection (*Id2*^{−/−}, *n* = 5; control, *n* = 3). **b**, The number of cytokine-producing/degranulating CXCR5⁺CD8⁺ T cells after stimulation with GP33-41 peptide and the expression of inhibitory molecules in CXCR5⁺CD8⁺ T cells in the spleens of control and *Id2*^{−/−} mice (*Id2*^{−/−}, *n* = 5 or 4; control, *n* = 3). **c**, Viral titers in the indicated tissues of control and *Id2*^{−/−} mice (*Id2*^{−/−}, *n* = 4; control, *n* = 3). **d**, **e**, Donor bone marrow cells from either *Id2*^{−/−} (CD45.2) or wild-type (CD45.1) mice were mixed with cells from wild-type (CD45.1) mice and transferred to lethally irradiated wild-type (CD45.1) mice. On day 25 after Cl13 infection, the frequency, degranulation, cytokine production and expression of inhibitory molecules of *Id2*^{−/−} or wild-type CXCR5⁺CD8⁺ T cells were analysed (*n* = 4 or 3). **f**, ChIP–qPCR of the binding of E2A to *Cxcr5* (chr9:44266001–44267000) loci (*n* = 3). **g**, me3H3K4 and me3H3K27 modifications on the *Cxcr5* loci (*n* = 3). The data are representative of two (**f**, **g**) or three (**a–c**, **e**) independent experiments, and were analysed by two-tailed unpaired *t*-test (**a–c**, **e–f**). Error bars (**a–c**, **e–g**) denote s.e.m. **P* < 0.05; ***P* < 0.01; ****P* < 0.001.

To examine whether the CXCR5⁺CD8⁺ T-cell population can be exploited for immunotherapy against chronic viral infection, we adoptively transferred this population at three intervals into recipients with severe Cl13 infection (Fig. 4a). On day 5 after the final cell transfer, we observed a 100–1,000-fold reduction in the viral load in mice transferred with the CD44^{hi}CXCR5⁺ subset compared to that in mice treated with the CD44^{hi}CXCR5[−] subset (Fig. 4b). Blockade of the PD-1–PD-L1 pathway has become an effective strategy to reinvigorate exhausted CD8 T cells for the improved control of chronic viral infection^{22–25}. We found that the combination of anti-PD-L1 treatment and adoptive transfer of the CXCR5⁺ subset synergistically inhibited viral replication (Fig. 4c, d). These data indicate that the CXCR5⁺CD8⁺ T-cell subset has a great potential for use as an effective immunotherapy to treat chronic infections.

Taken together, we propose that, during chronic viral infection, the Id2–E2A axis drives virus-specific CD8 T cells to differentiate into distinct CXCR5⁺ and CXCR5[−] subsets. CXCR5[−]CD8⁺ T cells reside in the T-cell zone and undergo severe exhaustion owing to the inhibitory microenvironment *in situ*. By contrast, the CXCR5⁺CD8⁺ T cells migrate into B-cell follicles, in which the less-inhibitory microenvironment prevents the rapid loss of effector functions in these cells. CXCR5⁺CD8⁺ T cells eventually turn into CXCR5[−]CD8⁺ T cells, accompanied by increased Id2 expression. The *de novo* converted CXCR5[−]CD8⁺ T cells possess better cytotoxicity and, when exiting B-cell follicles, they are capable of more effectively clearing virus-infected cells outside follicles (Extended Data Fig. 10). In chronic simian immunodeficiency virus or HIV infection, T_{FH} cells are productively infected in B-cell follicles, whereas very few CXCR5⁺CD8⁺ T cells are present in this privileged area, resulting in the persistent productive infection in B-cell follicles^{20,21,26,27}. According to our results, the adoptive transfer of HIV-specific CXCR5⁺CD8⁺ T cells or targeting

type (CD45.2) mice were mixed with cells from wild-type (CD45.1) mice and transferred to lethally irradiated wild-type (CD45.1) mice. On day 25 after Cl13 infection, the frequency, degranulation, cytokine production and expression of inhibitory molecules of *Id2*^{−/−} or wild-type CXCR5⁺CD8⁺ T cells were analysed (*n* = 4 or 3). **f**, ChIP–qPCR of the binding of E2A to *Cxcr5* (chr9:44266001–44267000) loci (*n* = 3). **g**, me3H3K4 and me3H3K27 modifications on the *Cxcr5* loci (*n* = 3). The data are representative of two (**f**, **g**) or three (**a–c**, **e**) independent experiments, and were analysed by two-tailed unpaired *t*-test (**a–c**, **e–f**). Error bars (**a–c**, **e–g**) denote s.e.m. **P* < 0.05; ***P* < 0.01; ****P* < 0.001.

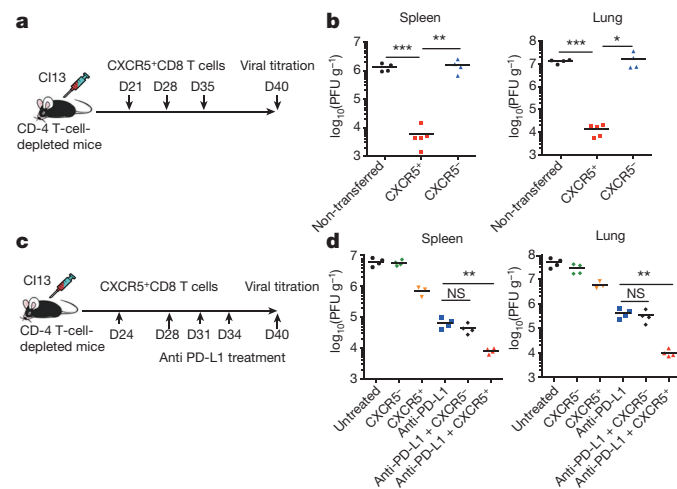


Figure 4 | CXCR5⁺CD8⁺ T cells exhibit greater therapeutic potential than CXCR5[−]CD8⁺ T cells in the control of chronic viral infection. **a**, **b**, Equal numbers of CD44^{hi}CXCR5⁺ or CD44^{hi}CXCR5[−]CD8⁺ T cells sorted from day 8 Cl13-infected mice were adoptively transferred into Cl13-infected CD4⁺ T-cell-depleted recipients on days 21, 28 and 35 after infection. Five days after the final transfer, virus titration was determined in indicated tissues (non-transferred, *n* = 4; CXCR5⁺, *n* = 5; CXCR5[−], *n* = 4). **c**, **d**, Cl13-infected CD4⁺ T-cell-depleted mice received equal numbers of CD44^{hi}CXCR5⁺ or CD44^{hi}CXCR5[−]CD8⁺ T cells sorted from day 8 Cl13-infected mice on day 21 after infection and were then treated with anti-PD-L1. Virus titration was performed 6 days after the final anti-PD-L1 treatment (CXCR5⁺, *n* = 3; other, *n* = 4). The data are representative of two independent experiments, and were analysed by two-tailed unpaired *t*-test (**b**, **d**). Error bars (**b**, **d**) denote s.e.m. **P* < 0.05; ***P* < 0.01; ****P* < 0.001. NS, not significant.

the E2A–Id2 axis in HIV-specific cytotoxic lymphocytes may potentially overcome the B-follicle sanctuary for more effectively purging HIV infection. Furthermore, given the largely shared mechanisms of T-cell exhaustion between chronic viral infection and cancer^{28–30}, this study may shed new light on cancer immunotherapy.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 December 2015; accepted 20 July 2016.

Published online 2 August 2016.

1. Zajac, A. J. *et al.* Viral immune evasion due to persistence of activated T cells without effector function. *J. Exp. Med.* **188**, 2205–2213 (1998).
2. Wherry, E. J. T cell exhaustion. *Nat. Immunol.* **12**, 492–499 (2011).
3. Gallimore, A. *et al.* Induction and exhaustion of lymphocytic choriomeningitis virus-specific cytotoxic T lymphocytes visualized using soluble tetrameric major histocompatibility complex class I-peptide complexes. *J. Exp. Med.* **187**, 1383–1393 (1998).
4. Schietinger, A. & Greenberg, P. D. Tolerance and exhaustion: defining mechanisms of T cell dysfunction. *Trends Immunol.* **35**, 51–60 (2014).
5. Speiser, D. E. *et al.* T cell differentiation in chronic infection and cancer: functional adaptation or exhaustion? *Nat. Rev. Immunol.* **14**, 768–774 (2014).
6. Deng, K. *et al.* Broad CTL response is required to clear latent HIV-1 due to dominance of escape mutations. *Nature* **517**, 381–385 (2015).
7. Wherry, E. J. & Kurachi, M. Molecular and cellular insights into T cell exhaustion. *Nat. Rev. Immunol.* **15**, 486–499 (2015).
8. Jones, R. B. & Walker, B. D. HIV-specific CD8⁺ T cells and HIV eradication. *J. Clin. Invest.* **126**, 455–463 (2016).
9. Zehn, D., Utzschneider, D. T. & Thimme, R. Immune-surveillance through exhausted effector T-cells. *Curr. Opin. Virol.* **16**, 49–54 (2016).
10. Cyster, J. G. *et al.* Follicular stromal cells and lymphocyte homing to follicles. *Immunol. Rev.* **176**, 181–193 (2000).
11. Crotty, S. T follicular helper cell differentiation, function, and roles in disease. *Immunity* **41**, 529–542 (2014).
12. Haynes, N. M. *et al.* Role of CXCR5 and CCR7 in follicular T_H cell positioning and appearance of a programmed cell death gene-1^{high} germinal center-associated subpopulation. *J. Immunol.* **179**, 5099–5108 (2007).
13. Kim, H. J., Verbinen, B., Tang, X., Lu, L. & Cantor, H. Inhibition of follicular T-helper cells by CD8⁺ regulatory T cells is essential for self tolerance. *Nature* **467**, 328–332 (2010).
14. Kim, H. J. & Cantor, H. Regulation of self-tolerance by Qa-1-restricted CD8⁺ regulatory T cells. *Semin. Immunol.* **23**, 446–452 (2011).
15. Belle, I. & Zhuang, Y. E proteins in lymphocyte development and lymphoid diseases. *Curr. Top. Dev. Biol.* **110**, 153–187 (2014).
16. Miyazaki, M. *et al.* The opposing roles of the transcription factor E2A and its antagonist Id3 that orchestrate and enforce the naive fate of T cells. *Nat. Immunol.* **12**, 992–1001 (2011).
17. Shaw, L. A. *et al.* Id2 reinforces T_H1 differentiation and inhibits E2A to repress T_{FH} differentiation. *Nat. Immunol.* **17**, 834–843 (2016).
18. Vezys, V. *et al.* Continuous recruitment of naive T cells contributes to heterogeneity of antiviral CD8 T cells during persistent infection. *J. Exp. Med.* **203**, 2263–2269 (2006).
19. Quigley, M. F., Gonzalez, V. D., Granath, A., Andersson, J. & Sandberg, J. K. CXCR5⁺ CCR7[−] CD8 T cells are early effector memory cells that infiltrate tonsil B cell follicles. *Eur. J. Immunol.* **37**, 3352–3362 (2007).
20. Connick, E. *et al.* CTL fail to accumulate at sites of HIV-1 replication in lymphoid tissue. *J. Immunol.* **178**, 6975–6983 (2007).
21. Banga, R. *et al.* PD-1⁺ and follicular helper T cells are responsible for persistent HIV-1 transcription in treated aviremic individuals. *Nat. Med.* **22**, 754–761 (2016).
22. Barber, D. L. *et al.* Restoring function in exhausted CD8 T cells during chronic viral infection. *Nature* **439**, 682–687 (2006).
23. Freeman, G. J., Wherry, E. J., Ahmed, R. & Sharpe, A. H. Reinvigorating exhausted HIV-specific T cells via PD-1-PD-1 ligand blockade. *J. Exp. Med.* **203**, 2223–2227 (2006).
24. Velu, V. *et al.* Enhancing SIV-specific immunity in vivo by PD-1 blockade. *Nature* **458**, 206–210 (2009).
25. Fuller, M. J. *et al.* Immunotherapy of chronic hepatitis C virus infection with antibodies against programmed cell death-1 (PD-1). *Proc. Natl Acad. Sci. USA* **110**, 15001–15006 (2013).
26. Fukazawa, Y. *et al.* B cell follicle sanctuary permits persistent productive simian immunodeficiency virus infection in elite controllers. *Nat. Med.* **21**, 132–139 (2015).
27. Miles, B. & Connick, E. T_{FH} in HIV latency and as sources of replication-competent virus. *Trends Microbiol.* **24**, 338–344 (2016).
28. Pauken, K. E. & Wherry, E. J. Overcoming T cell exhaustion in infection and cancer. *Trends Immunol.* **36**, 265–276 (2015).
29. Verdeil, G., Fuertes Marraco, S. A., Murray, T. & Speiser, D. E. From T cell “exhaustion” to anti-cancer immunity. *Biochim. Biophys. Acta* (2015).
30. Kim, P. S. & Ahmed, R. Features of responding T cells in cancer and chronic infection. *Curr. Opin. Immunol.* **22**, 223–230 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank R. Ahmed (Emory University) for providing the P14 TCR transgenic mice, retroviral vectors and LCMV Armstrong and Cl13 viruses; Y. Zhuang (Duke University) for providing the *Id2^{fl/fl}* mice; the core facility centre of Third Military Medical University for helping us with cell sorting; T. Wu (NIH) for insightful discussion. The work was supported by National Basic Research Program of China (973 program, 2013CB531500, to L.Y.; 2014CB542501 to H.Q.), the National Natural Science Foundation of China (81220108024 to Y.W.; 81471624 to L.Y.; U1202228 to J.X.; No. 81425011, 81330070 to H.Q.; No.31500733 to Q.B.).

Author Contributions L.Y. conceived the project. R.H., S.H., and C.L. performed both *in vivo* and *in vitro* experiments. Q.B., M.H., T.S., G.W., and T.N. performed RNA-seq and bioinformatics' analysis. A.Z., Y.Y., X.Y., L.X., X.C., Y.H., P.W., K.D., Y.C., J.O., Y.L., X.Z., and H.L. performed LCMV- and HIV-associated experiments; C.Z. performed thymectomy; X.Z. performed the reporter assay. L.Y., Y.W., H.Q., and J.X. designed the study, analysed the data and wrote the manuscript; L.Y., and Y.W. supervised the study.

Author Information RNA-seq data has been deposited in Gene Expression Omnibus under accession number GSE74148. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.W. (wuyuzhang@tmmu.edu.cn), H.Q. (qihai@mail.tsinghua.edu.cn) or L.Y. (yelilincmv@tmmu.edu.cn).

METHODS

Mice, virus and infections. The *Cd8*^{-/-}, *Cxcr5*^{-/-}, *Cd4*^{Cre} transgenic, μ MT and C57BL/6J (CD45.1 and CD45.2) mice were obtained from Jackson Laboratories. The *Id2*^{fl/fl} mice were a gift from Y. Zhuang, Duke University. The P14 (CD90.1) T-cell receptor (TCR) transgenic mice were from R. Ahmed, Emory University. The lymphocytic choriomeningitis virus (LCMV) Armstrong and clone 13 (Cl13) strains were gifts from R. Ahmed at Emory University. Mice were infected intraperitoneally (i.p.) with LCMV-Armstrong (2×10^5 plaque-forming units (PFU)) or intravenously (i.v.) with LCMV-Cl13 (2×10^6 PFU). Mice were infected at 6–10 weeks of age, and both sexes were included without randomization or blinding. No statistical method was used to predetermine sample size. The number of mice used in each experiment to reach statistical significance was determined on the basis of previous experience. Bone marrow chimaeras were infected after 8–10 weeks of reconstitution. Splenic chimaeras were infected 12–18 h after splenocyte transfer. Mice infected with LCMV were housed in accordance with institutional biosafety regulations of the Third Military Medical University. All mice were used in accordance with the guidelines of the Institutional Animal Care and Use Committees of the Third Military Medical University.

Generation of CXCR5–GFP knock-in mice. The CXCR5–GFP knock-in mice were generated by the insertion of an IRES–GFP construct after the open reading frame of *Cxcr5* by homologous recombination. A Neo cassette and diphtheria toxin were used as positive and negative selection markers in the targeting vector, respectively. Targeted embryonic stem clones were injected into C57BL/6J blastocysts to generate chimaeras. CXCR5–GFP knock-in reporter mice were obtained after deletion of Neo cassette by crossing with Cre-deleter mice. The generation of CXCR5–GFP knock-in mice were conducted by Beijing Biocytogen Co. Ltd.

Immunohistochemistry. Fresh spleens and lymph nodes were fixed with 1% paraformaldehyde for 10 h and subsequently dehydrated with 30% sucrose, followed by instant freezing in optimum cutting temperature compound. Sections 16 μ m in thickness were cut with a Leica Cryostat, mounted on Superfrost Plus glass slides. Staining reagents include eFluor450 anti-IgD (eBioscience), PE anti-CD8a (eBioscience), FITC anti-CD3e (Biolegend), APC anti-CD45.2 (eBioscience), rabbit monoclonal antibody against human CD20 (Abcam), mouse monoclonal antibody against human CD8 (Abcam), Alexa Fluor 488 anti-mouse IgG (Abcam), Alexa Fluor (R) 647 anti-rabbit IgG Fab2 (Cell Signaling). Images were acquired with an Olympus FV1000 or a Zeiss LSM 510 confocal fluorescence microscope using 20 \times air lens and were processed with Bitplane Imaris or LSM Image Examiner software (Zeiss).

Flow cytometry and antibodies. Mouse CXCR5 staining was performed in FACS buffer (PBS with 2% FBS) containing 1% BSA and 2% normal mouse serum. The cells were first stained with purified rat anti-mouse CXCR5 antibody (BD Bioscience) at 4°C for 1 h; then cells were washed and stained with biotin–streptavidin-conjugated goat anti-rat IgG (Jackson ImmunoResearch) on ice for 30 min; finally, cells were washed and stained with streptavidin (eBioscience) and other surface antibodies on ice for 30 min. Surface staining was performed in PBS containing 2% BSA or FBS (wt/vol). For intracellular cytokine production analysis, splenocytes were first stimulated by the GP33 (amino acid sequence: KAVYNFATC) or GP276 (amino acid sequence: SGVENPGGYCL) peptides (0.2 μ g ml⁻¹) and brefeldin A for 5 h at 37°C. Following surface staining, intracellular cytokine staining was performed with a Cytofix/Cytoperm Fixation/Permeabilization Kit (554714, BD Biosciences) according to the manufacturer's instructions. To detect degranulation, splenocytes were stimulated for 5 h in the presence of indicated peptide (0.2 μ g ml⁻¹), brefeldin A, anti-CD107a and anti-CD107b antibodies (BD Biosciences). The antibodies used for flow cytometry are listed in Supplementary Table 1. Major histocompatibility complex (MHC) class I peptide tetramers of H-2D^b complex with LCMV GP33-41 and GP276-286 were obtained from R. Ahmed (Emory University). The HLA pentamers were purchased from Proimmune. E2A (isoform E47) and Id2 staining was performed with a Foxp3 Staining Buffer Set (eBioscience) according to the manufacturer's instructions after surface staining. Samples were collected by using a FACSCanto (BD Bioscience) and analysed by FlowJo (Treestar).

Cell sorting and adoptive transfer. Cell sorting was performed on a FACSARIAII (BD Biosciences). The purity for all populations was >95%.

In each individual experiment, equal numbers (1×10^6) of CD44^{hi}CXCR5⁺ and CD44^{hi}CXCR5⁻ T cells were adoptively transferred to each recipient mouse intravenously. For P14 experiments, a total of 4,000 P14 cells were transferred into C57BL/6J mice that were infected with LCMV Cl13 on the following day.

In vivo killing assay. In vivo killing assay was performed as previously described³¹. Briefly, target cells from C57BL/6J (CD45.1) mice were labelled with CFSE (Life technologies) or Cell-trace Violet (Life technologies) at either 100 nM or 1 μ M. The labelled cells were then pulsed with 2 μ g of LCMV-GP33-41 or GP276-286 peptides for 1 h at 37°C and then rinsed three times in RPMI 1640 with 10%

FBS. The peptide pulsed target cells were mixed with sorted CD44^{hi}CXCR5⁺ and CD44^{hi}CXCR5⁻ CD8 T cells at a 1:2 effector:target ratio (E:T) and transferred into naive C57BL/6J (CD45.2) mice. Mice were killed for analysis 5 h later. The E:T ratio was determined by normalizing all populations to the number of D^bGP33- or D^bGP276-tetramer-positive CD8⁺ T cells. The killing efficiency was determined as follows: $100 - ((\% \text{ peptide pulsed in infected} / \% \text{ un-pulsed in infected}) / (\% \text{ peptide pulsed in uninfected} / \% \text{ un-pulsed in uninfected})) \times 100$.

Ex vivo killing assay. Target cells from C57BL/6J (CD45.1) mice were labelled with Cell-trace Violet (Life technologies) at either 100 nM or 1 μ M. The labelled cells were then pulsed with 2 μ g of indicated peptides for 1 h at 37°C and then rinsed three times in RPMI 1,640 with 10% FCS. The peptide pulsed target cells were mixed with sorted GFP⁺ and GFP⁻ CD8 T cells at a 4:1 E:T ratio and co-cultured at 37°C for 5 h. The E:T ratio was determined by normalizing all populations to the number of D^bGP33 tetramer-positive CD8⁺ T cells. The killing efficiency was determined as follows: $100 - ((\% \text{ peptide pulsed in infected} / \% \text{ un-pulsed in infected}) / (\% \text{ peptide pulsed in uninfected} / \% \text{ un-pulsed in uninfected})) \times 100$.

Quantitative PCR. Cells were sorted on a FACSARIA (BD Biosciences) and RNA was extracted in Trizol LS reagent (Life Technologies) and reverse-transcribed using RevertAid Minus First Strand cDNA Synthesis Kit (Thermo Scientific). Relative quantification PCR (qPCR) was performed with QuantiFast SYBR Green PCR Kit (Qiagen) on a CFX96 Touch Real-Time System (Bio-Rad). Primer pairs for detection of mouse *Id2* and *E2A* and internal HPRT control are as follows: *Id2* (forward, 5'-CATCAGCATCTGTCTTGC-3'; reverse, 5'-GTGTTCTCCTGGTGAAATGG-3'), *E47* (forward, 5'-CAGCAGTGACCAGAACAG-3'; reverse, 5'-AAGGTGGCATAGGATTC-3') and HPRT (forward, 5'-GCGTCGTGATTAGCGATGATG-3'; reverse, 5'-CTCGAGCAAGTCTTTCAGTCC-3').

Bone marrow chimaera. Bone marrow cells from C57BL/6J (CD45.2) or *Cd4*^{Cre}-*Id2*^{fl/fl} (CD45.2) and bone marrow cells from C57BL/6J (CD45.1) mice were mixed and adoptively transferred intravenously at a 3:7 ratio into lethally irradiated (two doses of 550 rad each) wild-type C57BL/6J (CD45.1) mice. A total of 5 million bone marrow cells were transferred per mouse. Recipient mice were fed antibiotics for 2 weeks and allowed to reconstitute for at least 8 weeks before infection.

Splenic chimaeras. Total splenocytes from *Cd8*^{-/-} mice and from *Cxcr5*^{-/-} or wild-type mice were mixed and adoptively transferred intravenously at a 4:6 ratio into irradiated (600 rad) naive *Cd8*^{-/-} mice. A total of 50 million lymphocytes were transferred per mouse. LCMV infection was done 12–18 h after the cell transfer.

ELISA. LCMV-specific serum antibody titers were determined by ELISA as previously described³², using horseradish peroxidase (HRP)-conjugated goat anti-mouse IgG secondary antibodies (Southern Biotech).

Virus titration. The LCMV viral loads in tissue samples were quantified by a qPCR assay as described previously³³.

RNA-seq library construction. The total RNA from sorted CD44^{hi}CXCR5⁺ and CD44^{hi}CXCR5⁻ CD8⁺ T cells were extracted by Trizol reagent (Life Technologies), and then purified with Dnase I (Qiagen) treatment. The RNA-seq library construction for the RNA samples was according to the strand-specific RNA sequencing library preparation protocol³⁴. The mRNA transcripts were enriched by two rounds of poly (A+) selection with Dynabeads oligonucleotide (dT) 25 (Invitrogen) before library construction. The prepared library was sequenced with the Illumina HiSeq 2000 sequencer.

Bioinformatic analysis. The raw sequence reads were first aligned to mouse UniGene with bowtie (version 1.0.0) to estimate the insert-fragment size and the standard deviations, which are needed by TopHat2, were used to align the reads to the genome. Then TopHat2 was used to align the reads to the reference mouse genome (GRCm38) with the aligning parameter –bowtie1 and Ensemble annotated transcripts (version 77) as guide reference. The uniquely mapped reads were used for quantifying gene expression and differential gene expression evaluation was analyzed by Cuffdiff, a subpackage of Cufflinks (version 2.1.1) with Ensemble annotated genes (version 77).

Abundance of transcripts (including mRNAs, pseudogenes, non-coding RNAs and other predicted RNAs) were calculated and normalized in RPKM as described above from the raw RNA-seq data and used for Gene Set Enrichment Analysis (GSEA, Broad Institute)³⁵.

Retroviral constructs and transduction. MIGR1 (MSV-IRES-GFP) retroviral construct expressing E2A and MIT (MSCV-IRES-Thy1.1) retroviral construct expressing *Id2* were obtained from R. Ahmed. The self-inactivated retroviral reporter vector was modified as previously described³⁶. We first inserted the SV40 promoter into the modified construct. Then, we cloned the wild-type and mutant of *Cxcr5* intron regulatory regions (*Cxcr5*, +10,465 to +10,923) and inserted these sequences into the construct. The mutations are indicated in Extended Data Fig. 7e. All sequences were verified by sequencing. Retroviruses were packaged by transfection of 293T cells with the retroviral vectors along with the pCL^{eco} plasmid. Naive CD8⁺ T cells were isolated and purified from

naive C57BL/6J mice and were stimulated for 48 h with anti-CD3 ($0.2 \mu\text{g ml}^{-1}$; 17A2 Biolegend) and anti-CD28 ($0.5 \mu\text{g ml}^{-1}$; 37.51 Biolegend). P14 CD8⁺ T cells were activated *in vivo* by injection of 200 μg GP33-41 peptide into P14 transgenic mice. Eighteen hours later, activated CD8⁺ T cells were isolated and purified, and were spin-infected by centrifugation (800g) with freshly collected retrovirus supernatants, $8 \mu\text{g ml}^{-1}$ polybrene (Sigma-Aldrich) and 20 ng ml^{-1} of IL-2 (Miltenyi Biotec) at 37°C for 90 min. Then, CD8⁺ T cells were cultured for three days before analysis (in Extended Data Fig. 7f) or were transferred into recipient mice, followed by infection of the recipients with LCMV Cl13 (in Extended Data Fig. 7g, h).

Chromatin immunoprecipitation. The sorted CD44^{hi}CXCR5⁺ and CD44^{hi}CXCR5⁺CD8⁺ T cells were crosslinked for 10 min with 1% formaldehyde in medium. Chromatin fragments were prepared as previously described³⁷ and immunoprecipitated with antibody against E2A (sc-349X, Santa Cruz Biotechnology), me3H3K4 (CS 200580, Millipore), me3H3K27 (CS 200603, Millipore) or rabbit IgG (PP64B, Millipore) coupled with Dynabeads Protein G (Life Technologies). DNA was purified using a PCR purification kit (Qiagen) and eluted by water. qPCR was performed to quantitatively determine DNA segments by using the primers (*Cxcr5* forward, 5'-GACAGGGTGCCTGTTTTCAT-3'; reverse, 5'-TTCGGGTGTAATTGGTTTTTG-3') that flank putative E2A binding sites. The relative enrichment for the segment was calculated as firstly normalized to control IgG, followed by normalization to input DNA. The input DNA was defined as an aliquot of sheared chromatin before immunoprecipitation, and was used to normalize the sample to the amount of chromatin added to each ChIP.

In vivo antibody blockade. For PD-L1 blockade, 200 μg of rat anti-mouse PD-L1 antibody (10F9G2 BioXcell) were administered (i.p.) 3 times, every 3 days. For depletion of CD4⁺ T cells, mice were given 500 μg of anti-mouse CD4 antibody (GK1.5 BioXcell) (i.p.) on day -1 and day 1 after LCMV Cl-13 infection.

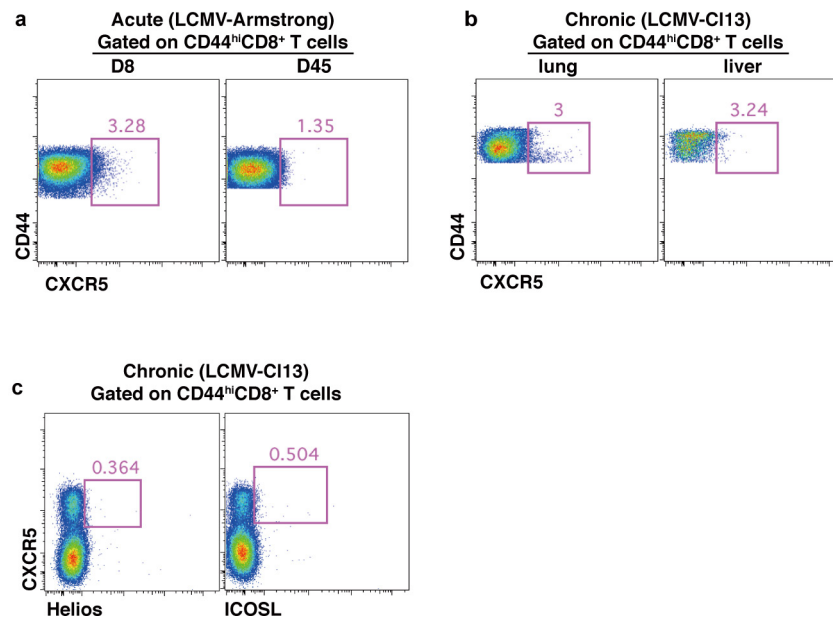
Human study subjects. Peripheral blood for the isolation of peripheral blood mononuclear cells (PBMCs) and lymph nodes were obtained from HIV-infected patients and HIV-negative donors. PBMCs were isolated with Ficoll (Sigma-Aldrich) gradient separation. No statistical method was used to predetermine

sample size. The detection limit for HIV-1 RNA is 50 copies per ml in the serum. The study was reviewed and approved by the Ethics Committee of Shanghai Public Health Clinical Center, Fudan University. Written informed consents were provided by all study participants. The detailed information of blood donors for viral load correlation analysis was listed in Supplementary Table 2.

Stimulation of HIV-specific CD8⁺ T cells. Overlapping sets of peptides covering HIV-1 pol, gag and env antigens (PepMix ULTRA Peptide Pools; JPT, Germany) were used to stimulate HIV-specific CD8⁺ T cells isolated from lymph nodes. The total cells isolated from lymph nodes were stimulated by the peptide pools ($1 \mu\text{g ml}^{-1}$, 100 μl per sample) and brefeldin A for 8 h at 37°C before surface and intracellular staining. The CD8⁺ T cells producing IFN- γ upon stimulation were defined as HIV-specific.

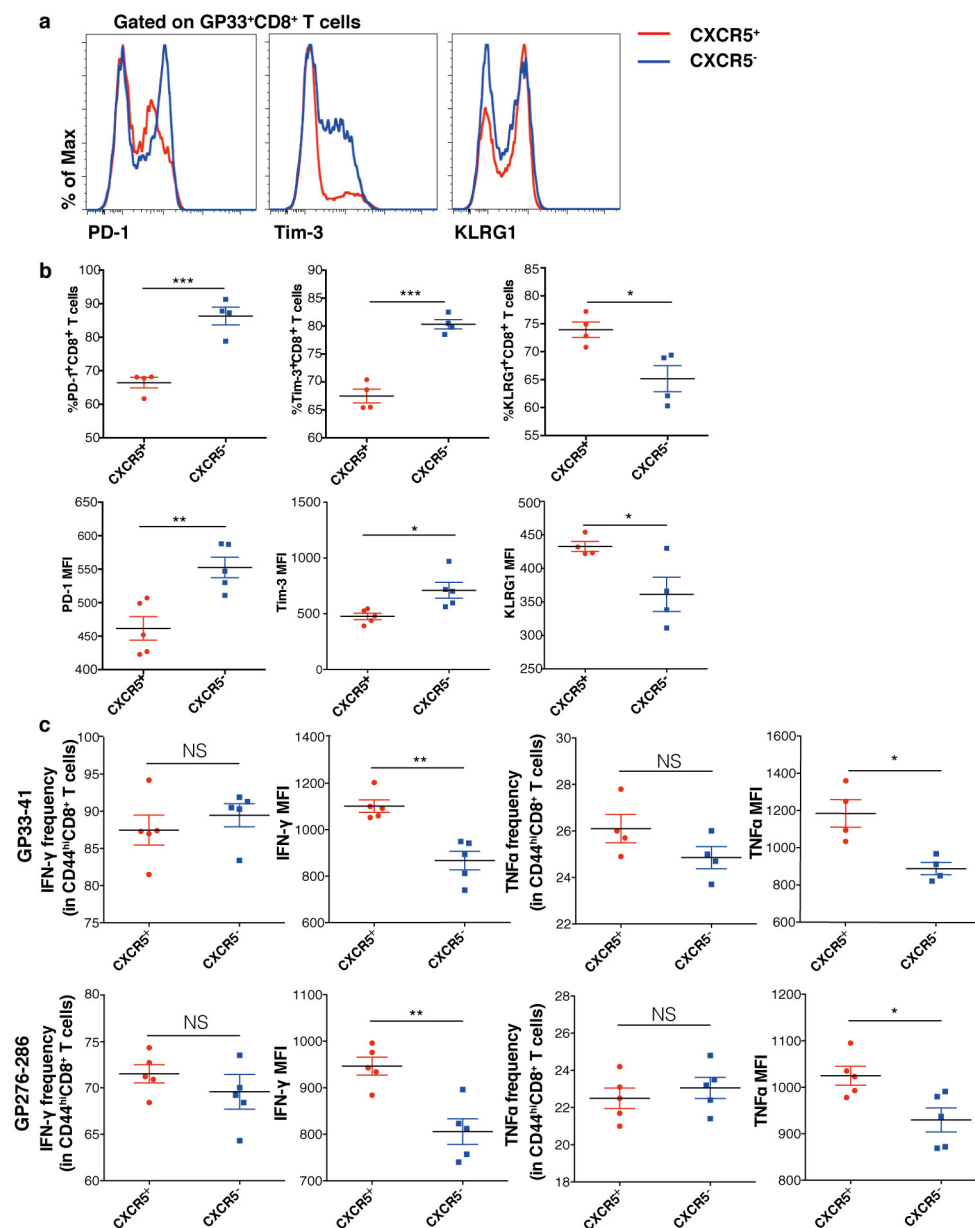
Statistical analysis. The statistical analysis was conducted with Prism 6.0 (GraphPad). A two-tailed unpaired Student *t* test with 95% confidence interval was used to calculate *P* values. For Extended Data Fig. 9b, f, g, a paired two-tailed *t*-test with 95% confidence interval was used for calculation of *P* values. For Extended Data Fig. 9c, a two-tailed nonparametric Spearman correlation test with 95% confidence interval was used for calculation of *r* and *P* values.

31. Barber, D. L., Wherry, E. J. & Ahmed, R. Cutting edge: rapid *in vivo* killing by memory CD8 T cells. *J. Immunol.* **171**, 27–31 (2003).
32. Rasheed, M. A. *et al.* Interleukin-21 is a critical cytokine for the generation of virus-specific long-lived plasma cells. *J. Virol.* **87**, 7737–7746 (2013).
33. McCausland, M. M. & Crotty, S. Quantitative PCR technique for detecting lymphocytic choriomeningitis virus *in vivo*. *J. Virol. Methods* **147**, 167–176 (2008).
34. Zhong, S. *et al.* High-throughput illumina strand-specific RNA sequencing library preparation. *Cold Spring Harb. Protoc.* **2011**, 940–949 (2011).
35. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
36. Zhou, X. *et al.* Differentiation and persistence of memory CD8⁺ T cells depend on T cell factor 1. *Immunity* **33**, 229–240 (2010).
37. Xu, L. *et al.* The transcription factor TCF-1 initiates the differentiation of T_{FH} cells during acute viral infection. *Nat. Immunol.* **16**, 991–999 (2015).



Extended Data Figure 1 | Virus-specific CXCR5⁺ CD8⁺ T cells are not apparent in acutely infected mice and in the non-lymphoid tissues of chronically infected mice and are not Qa-1-restricted. a, CXCR5 expression in virus-activated CD8⁺ T cells in the spleens of

Arm⁺-infected mice. **b,** CXCR5 expression in virus-activated CD8⁺ T cells in the lungs and livers of Cl13-infected mice. **c,** Helios and ICOSL expression in virus-activated CXCR5⁺ CD8⁺ T cells during Cl13 infection.

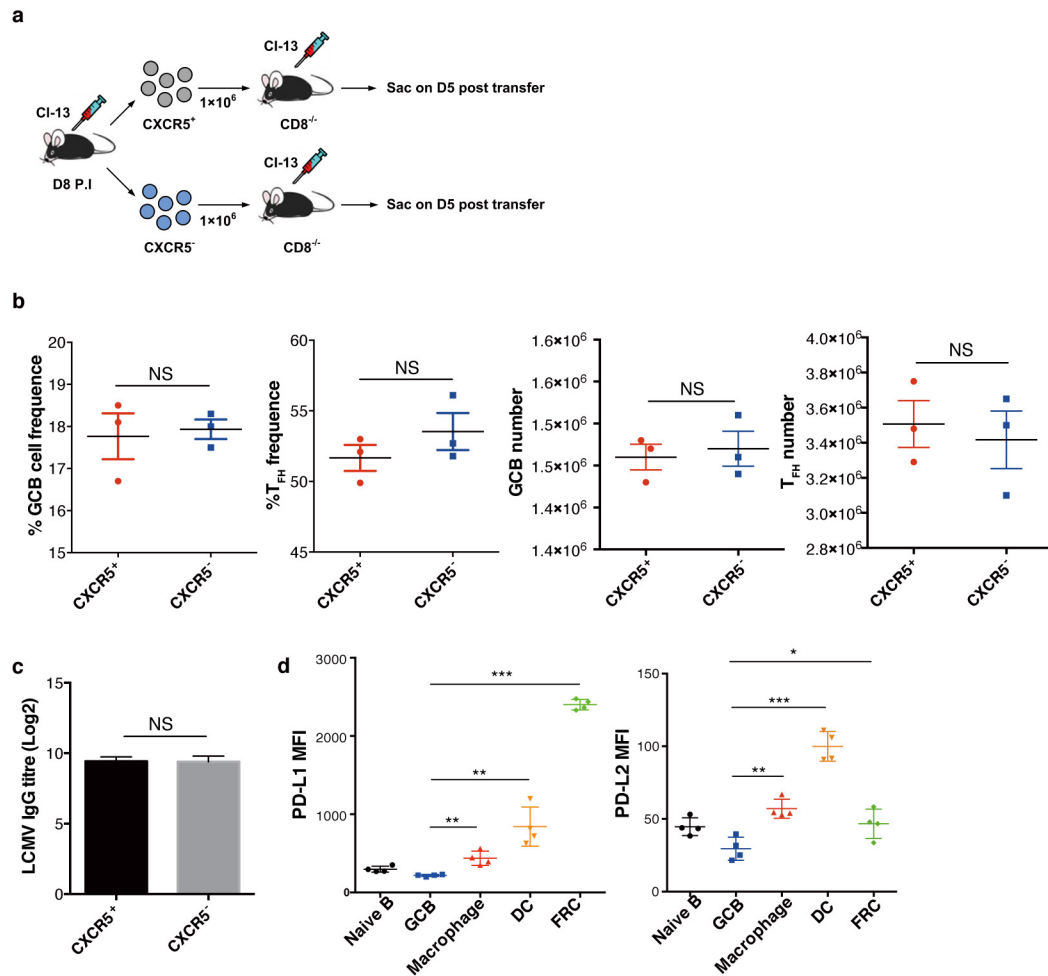


Extended Data Figure 2 | Virus-specific CXCR5⁺CD8⁺ T cells are less exhausted than CXCR5⁻CD8⁺ T cells on day 8 after CI13 infection.

a, b, PD-1, Tim-3 and KLRG1 expression on virus-specific CXCR5⁺ and CXCR5⁻CD8⁺ T cells in the spleens of CI13-infected mice on day 8 after infection ($n = 4$ or 5). MFI, mean fluorescence intensity.

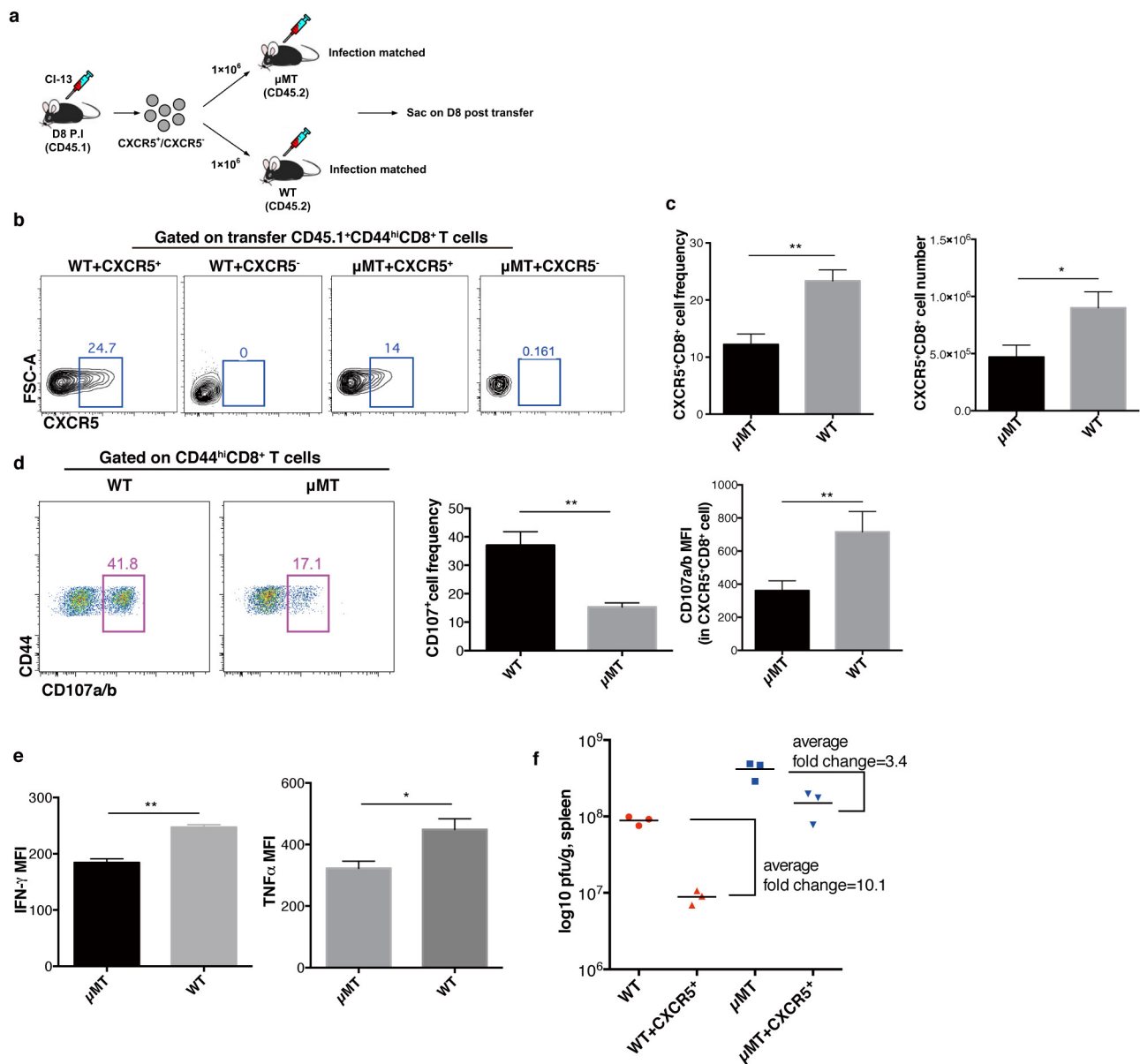
c, Upon stimulation with the indicated peptides, the cytokine production

of CXCR5⁺ and CXCR5⁻CD8⁺ T cells in the spleens of LCMV-CI13-infected mice was analysed on day 8 post-infection ($n = 4$ or 5). Data are representative of three independent experiments, and were analysed by two-tailed unpaired t -test (**b, c**). Error bars (**b, c**) denote s.e.m. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. NS, not significant.



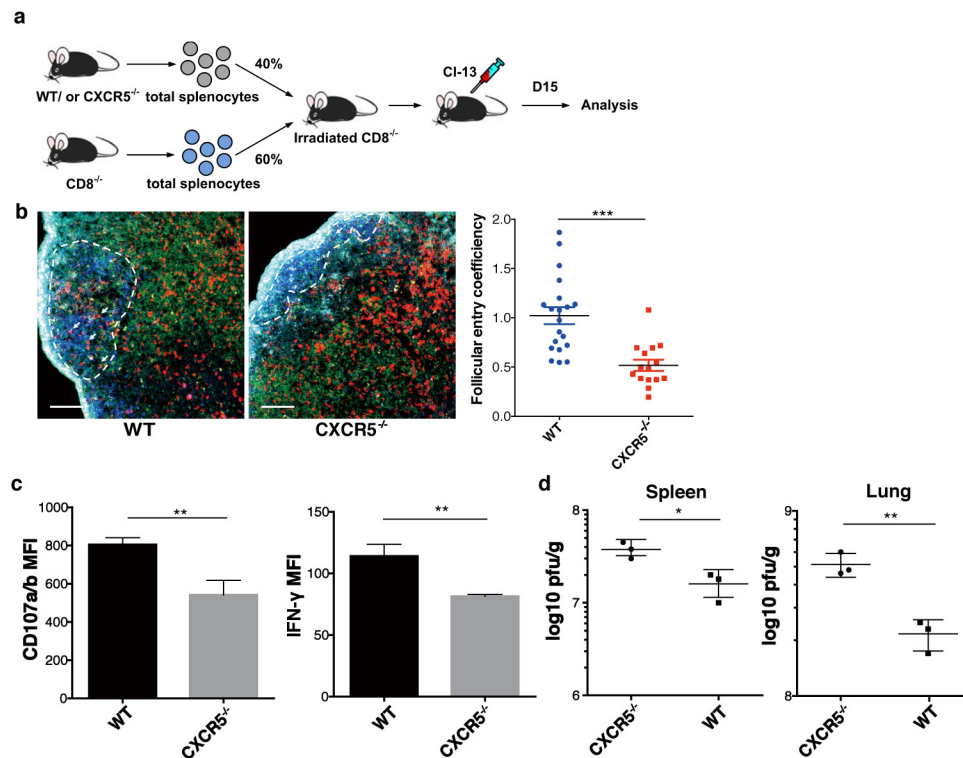
Extended Data Figure 3 | Virus-specific CXCR5⁺CD8⁺ T cells localized in B-cell follicles have minimal effect on germinal centre B and T_{FH} responses. **a, b**, Equal numbers of CXCR5⁺ and CXCR5⁻CD8⁺ T cells sorted from Cl13-infected mice were adoptively transferred into infection-matched CD8^{-/-} mice. On day 5 after transfer, frequency and number of germinal centre B cells and T_{FH} cells in the spleens of recipient mice were analysed ($n = 3$). **c**, Titration of LCMV-specific IgG in the serum of

recipient mice ($n = 3$). **d**, The expression levels of PD-L1 and PD-L2 on cell subsets residing in the T-cell zone and in B-cell follicles ($n = 4$). DC, dendritic cell; FRC, fibroblast reticular cell. The data are representative of three independent experiments, and were analysed by two-tailed unpaired *t*-test (**b–d**). Error bars (**b–d**) denote s.e.m. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. NS, not significant.



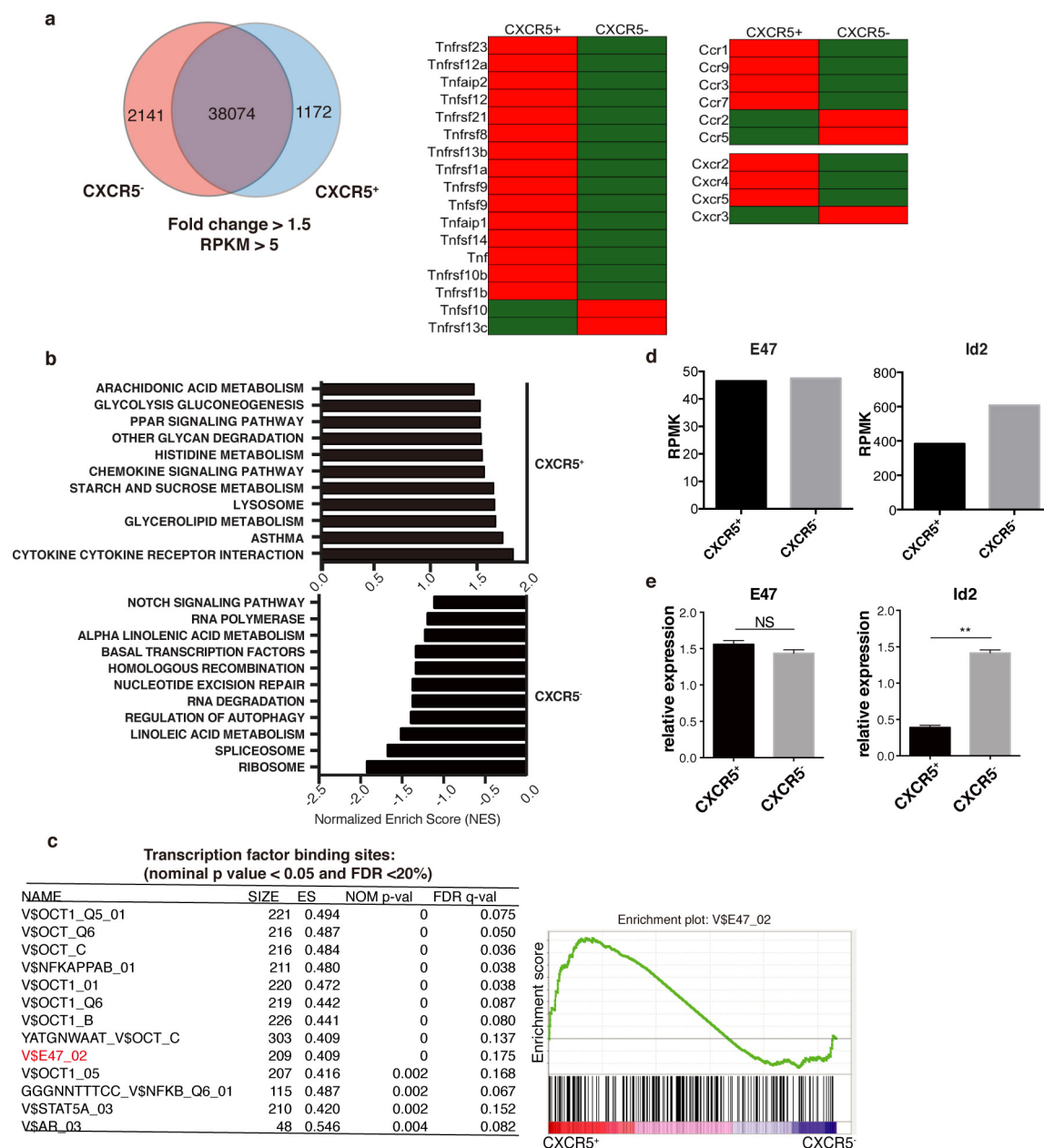
Extended Data Figure 4 | The maintenance of functional CXCR5⁺CD8⁺ T cells is dependent on follicle structures. **a**, Equal numbers of virus-activated CXCR5⁺CD8⁺ T and CXCR5⁻CD8⁺ T cells obtained from CI13-infected C57BL/6J (CD45.1) mice were adoptively transferred into infection-matched μMT (CD45.2) or C57BL/6 (CD45.2) (wild-type) mice. Analysis was performed on day 8 after transfer. **b**, **c**, Frequency and number of CD45.1⁺CXCR5⁺CD8⁺ T cells in the recipient mice ($n = 3$). **d**, **e**, On stimulation of peptide, surface CD107 expression and cytokine

production of CD45.1⁺CXCR5⁺CD8⁺ T cells in the recipient mice ($n = 3$). **f**, Viral titers in the indicated tissues obtained from control wild-type and μMT mice without cell transfer and from wild-type and μMT mice receiving CXCR5⁺CD8⁺ T cell transfer ($n = 3$). The data are representative of three independent experiments, and were analysed by two-tailed unpaired t -test (**c–e**). Error bars (**c–e**) denote s.e.m. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.



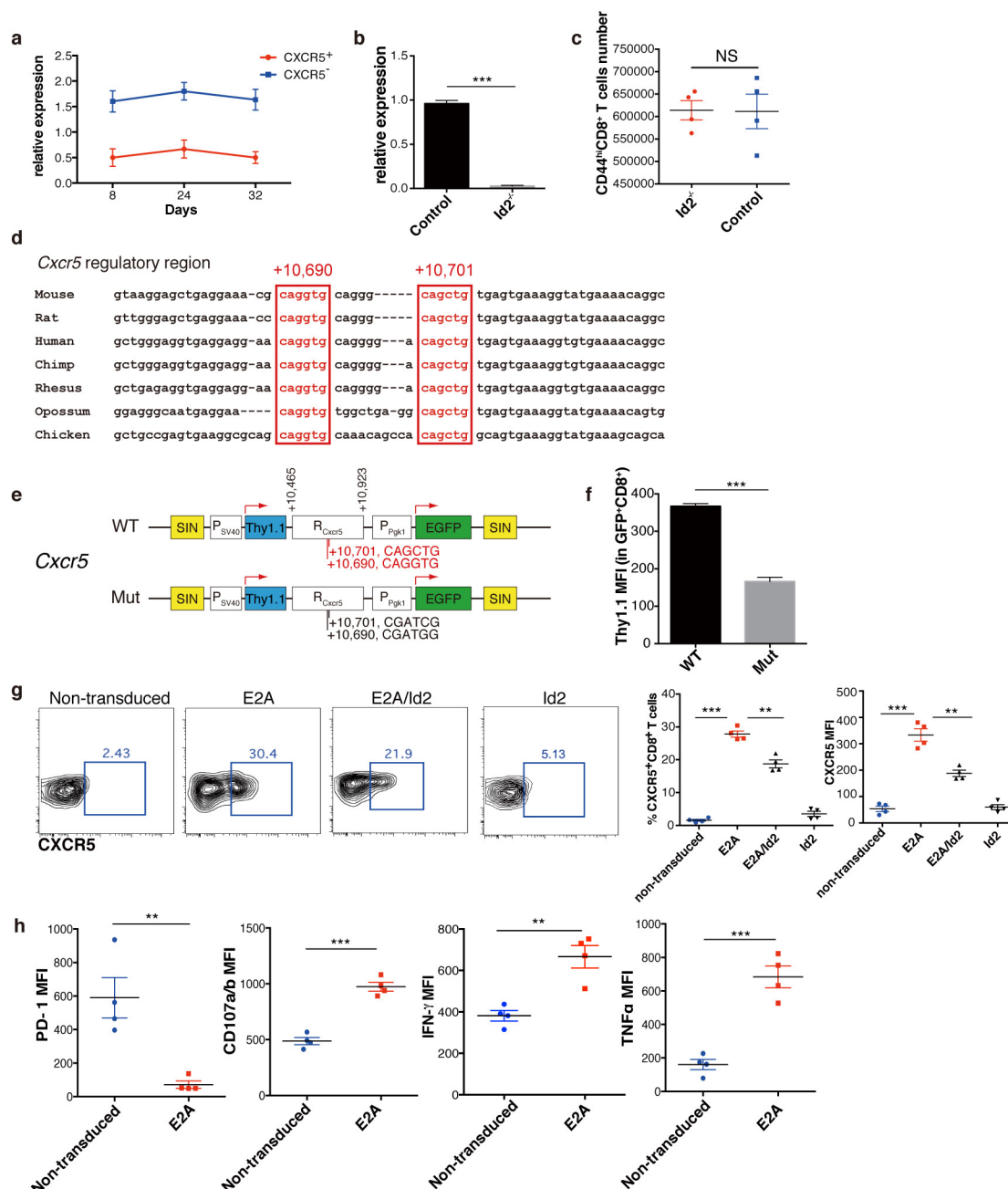
Extended Data Figure 5 | CXCR5 expression is critical for the localization of virus-activated CD8⁺ T cells to B-cell follicles. **a**, Set-up of splenic chimaera mice. Total splenocytes obtained from *Cxcr5*^{-/-} or wild-type mice were mixed with splenocytes obtained from *Cd8*^{-/-} mice and then transferred to non-lethally irradiated *Cd8*^{-/-} recipients and immediately infected with CI13. Analysis was performed on day 15 after infection. **b**, The localization of virus-activated CD8⁺ T cells in the lymph nodes was detected by confocal microscopy on day 15 after infection

(blue, IgD; red, CD8; green, CD3) and follicular entry coefficient was calculated (*Cxcr5*^{-/-}, *n* = 15; wild-type, *n* = 20). Scale bar, 100 μm. **c**, The CD107 expression and IFN-γ secretion of wild-type and *Cxcr5*^{-/-} CD8⁺ T cells upon peptide stimulation (*n* = 3). **d**, Viral titers in the indicated tissues from mice that received splenocytes from *Cxcr5*^{-/-} or wild-type mice (*n* = 3). Data are representative of three independent experiments, and were analysed by two-tailed unpaired *t*-test (**b–d**). Error bars (**b–d**) denote s.e.m. **P* < 0.05; ***P* < 0.01; ****P* < 0.001.



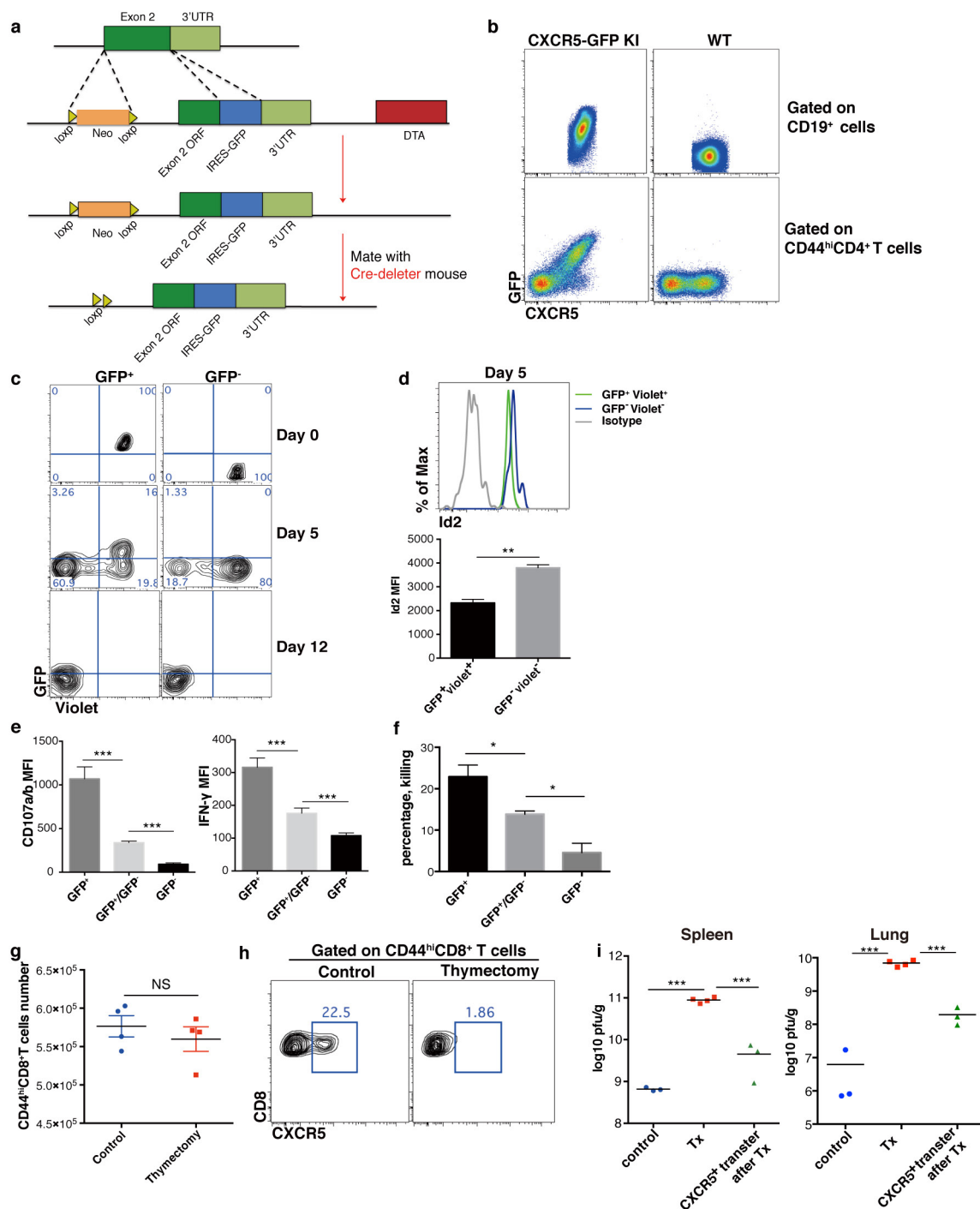
Extended Data Figure 6 | Distinct transcriptional profiles of CXCR5⁺ and CXCR5⁻ CD8⁺ T-cell populations. **a**, Transcriptomic profiling of CXCR5⁺ and CXCR5⁻ cell subsets. **b**, Gene Ontology (GO) enrichment was analysed using Gene Set Enrichment Analysis (GSEA) and significantly enriched (P value < 0.05) molecular function GO terms were shown with their enrichment scores. **c**, The enrichment of gene sets containing genes sharing upstream *cis*-regulatory motifs of transcription factor binding sites were assessed using GSEA. The transcription factor binding sites with significant enrichment (P value < 0.05) in CXCR5⁺CD8⁺ cells were listed (left). The GSEA result of the gene set including the E47 (E2A isoform) binding site (denoted as V\$E47_02 in

the Molecular Signatures Database version 3.0) was shown (right). **d**, The normalized expression levels of Id2 and E2A isoform E47 in CXCR5⁺ and CXCR5⁻CD8⁺ cells were calculated on the basis of RNA-seq data and was expressed in reads per kilobase per million mapped reads. **e**, qPCR analysis of the expression levels of Id2 and E2A isoform E47 in CXCR5⁺ and CXCR5⁻CD8⁺ cells. Data are from one experiment with two biological replicates (**a–d**) or are representative of three independent experiments (**e**), and were analysed by two-tailed unpaired *t*-test (**e**). Error bars (**e**) denote s.e.m. ** P < 0.01. NS, not significant.



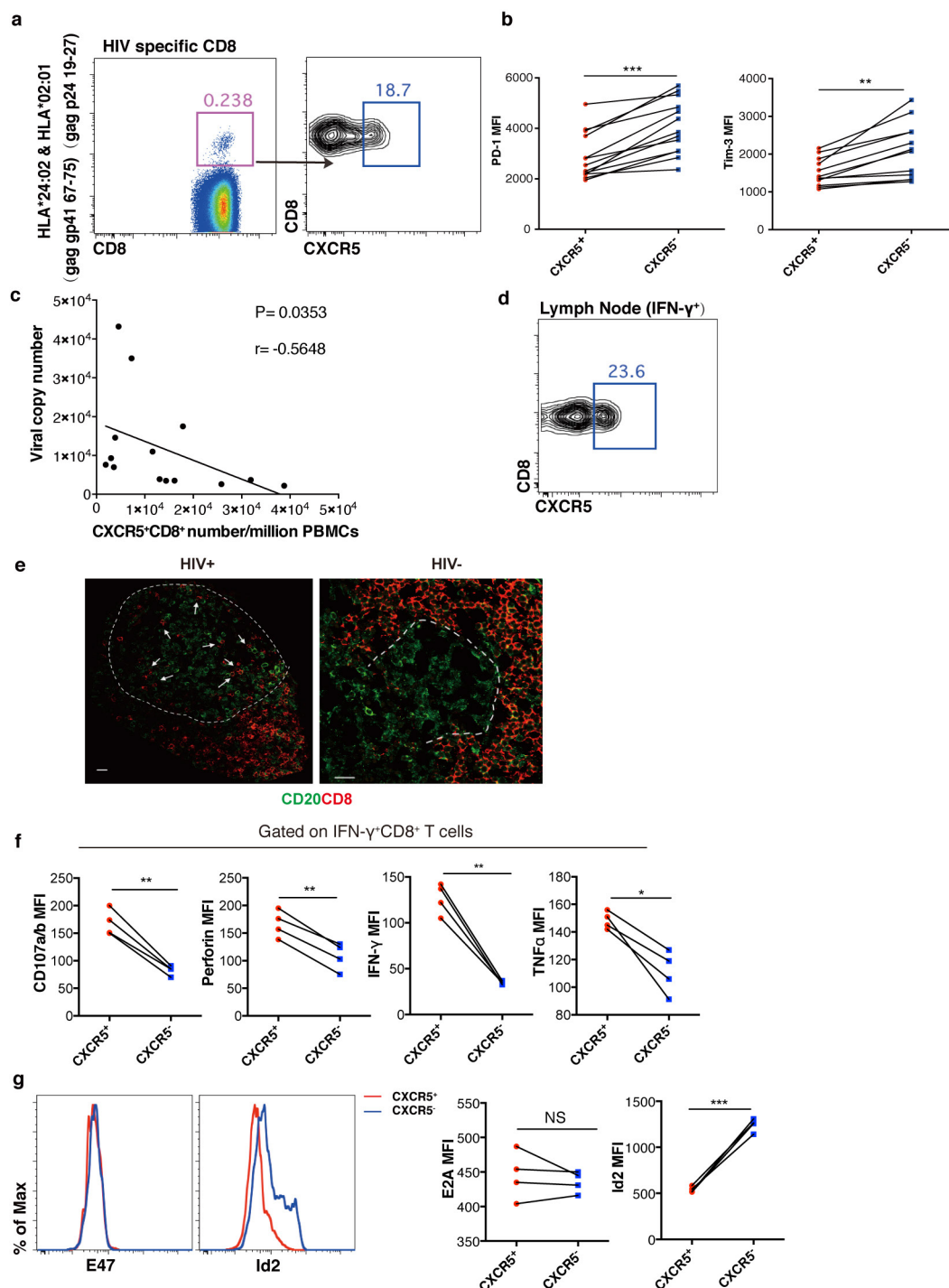
Extended Data Figure 7 | E2A regulates the transcription of *Cxcr5* by directly binding to DNA loci. **a**, Kinetic analysis of Id2 expression levels in CXCR5⁺ and CXCR5⁻ CD8⁺ T cells during C113 infection by qPCR ($n = 3$). **b**, Id2 mRNA expression in CD8⁺ T cells in the spleens of littermate control (control) and *Id2*^{-/-} mice ($n = 3$). **c**, The number of CD44^{hi}CD8⁺ T cells in the spleens of control and *Id2*^{-/-} mice on day 25 after C113 infection ($n = 4$). **d**, An alignment of putative E2A-binding sites in the *Cxcr5* intron. The conserved E2A-binding motif 'CASSTG' (or 'GTSSAC' on the reverse strand) is highlighted in red, and its locations relative to the transcriptional start site (TSS) of *Cxcr5* are marked. **e**, Retroviral reporter constructs containing a wild-type or mutated *Cxcr5* regulatory region and the Psv40 promoter, as well as self-inactivating mutations in the long terminal repeats (SIN), a sequence encoding Thy-1.1, and a PGK-EGFP cassette (including P-Pgk1

(a promoter of the gene encoding phosphoglycerate kinase 1) and EGFP). Arrows indicate the transcription start site and orientation, and the numbers shown above indicate the position. **f**, Thy-1.1 expression levels on GFP⁺CD8⁺ T cells transduced with a reporter construct containing wild-type or mutated *Cxcr5* regulatory region, MFI of Thy-1.1 was normalized to GFP expression ($n = 3$). **g**, CXCR5 expression in non-transduced, E2A-overexpressing, Id2-E2A-co-overexpressing and Id2-overexpressing P14 CD8⁺ T cells on day 8 after C113 infection ($n = 4$). E2A refers to E47 isoform. **h**, PD-1 and CD107a/b surface expression levels and cytokine production in non-transduced P14 cells and E2A-overexpressing P14 cells ($n = 4$). Data are representative of three independent experiments, and were analysed by two-tailed unpaired *t*-test (a–c, f–h). Error bars (a–c, f–h) denote s.e.m. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. NS, not significant.



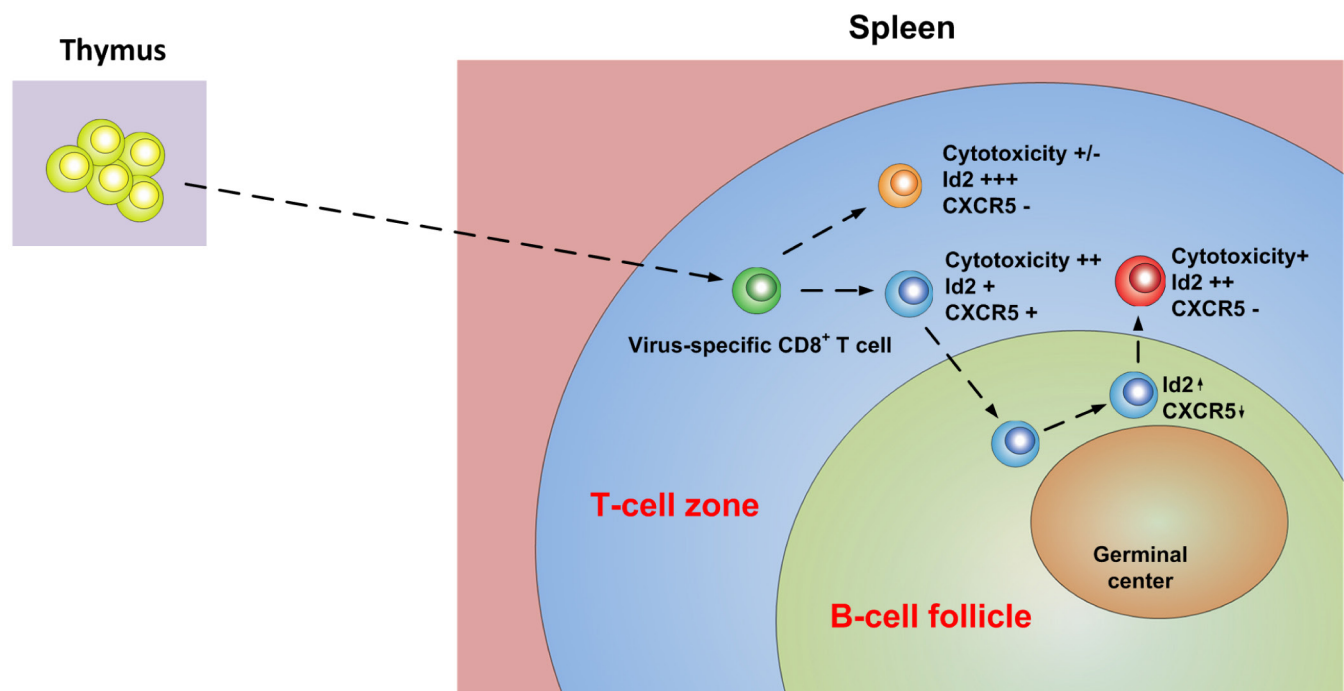
Extended Data Figure 8 | Virus-activated CXCR5⁺CD8⁺ T cells are converted into CXCR5⁻CD8⁺ T cells. **a**, Schematic map showing the construction of CXCR5-GFP knock-in mice. **b**, CXCR5-staining and GFP expression in CD19⁺ cells and in CD44^{hi}CD4⁺ T cells in CXCR5-GFP knock-in mice and from wild-type mice. **c**, GFP⁺CD44^{hi}CD8⁺ T cells and GFP⁻CD44^{hi}CD8⁺ T cells were sorted from day 8 Cl13-infected CXCR5-GFP knock-in mice (CD45.2). The cells were labelled with Celltrace Violet and then transferred into infection-matched wild-type recipients (CD45.1). The presence of GFP and Celltrace Violet in the transferred cells (CD45.2) was detected on days 0, 5, and 12 after transfer. **d**, Id2 expression levels in GFP⁺Violet^{hi}CD8⁺ T cells and in GFP⁻Violet^{lo}CD8⁺ T cells from recipient mice receiving GFP⁺CD8⁺ T cells transfer on day 5 after transfer ($n = 3$). **e**, Surface expression of CD107 and IFN- γ production in GFP⁺CD8⁺ T cells, newly converted GFP⁻CD8⁺ (GFP⁺/GFP⁻)

T cells and GFP⁻CD8⁺ T cells (GFP⁻, $n = 4$, GFP⁺/GFP⁻ and GFP⁺, $n = 3$). **f**, Equal numbers of GFP⁺CD8⁺ T cells, GFP⁺/GFP⁻ T cells and GFP⁻CD8⁺ T cells were co-cultured with peptide-coated target cells *ex vivo*, respectively. Five hours later, the killing efficiency of the effector cells was analysed ($n = 3$). **g**, **h**, The number of CD44^{hi}CD8⁺ T cells and the frequency of CXCR5⁺CD8⁺ T cells in the spleens of control mice infected on day 28 and thymectomized mice (subject to the surgery at day 21 after infection) ($n = 4$). **i**, Viral titers in the indicated tissues of control mice, mice received thymectomy and mice received CXCR5⁺CD8⁺ T cell transfer after thymectomy (control and CXCR5⁺ transfer, $n = 3$; thymectomy, $n = 4$). Data are representative of three independent experiments, and were analysed by two-tailed unpaired *t*-test (**d–g**, **i**). Error bars (**d–g**, **i**) denote s.e.m. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. NS, not significant.



Extended Data Figure 9 | The HIV-specific CXCR5⁺CD8⁺ T-cell subset is present in chronically HIV-infected patients. **a**, CXCR5 expression in HIV-specific CD8⁺ T cells in blood of HIV-infected patients. **b**, The expression levels of PD-1 and Tim-3 in HIV-specific CXCR5⁺ and CXCR5⁻CD8⁺ T cells in blood of HIV-infected patients (PD-1, $n = 13$; Tim-3, $n = 12$). **c**, The correlation between viral copy number in serum and CXCR5⁺CD8⁺ T cell number in blood in chronic HIV-infected patients prior to anti-retroviral treatment ($n = 14$). **d**, HIV-specific (IFN- γ ⁺) CXCR5⁺CD8⁺ T cells in lymph nodes of HIV-infected patients. **e**, CD8⁺ T-cell localization in the lymph nodes of HIV-infected patients and HIV-

negative donors by confocal microscopy (green, CD20; red, CD8). Scale bar, 20 μ m. **f**, The expression levels of CD107 and perforin and cytokine production in HIV-specific CXCR5⁺ and CXCR5⁻CD8⁺ T cells in lymph nodes of HIV-infected patients ($n = 4$). **g**, The expression levels of E2A isoform E47 and Id2 in IFN- γ ⁺CXCR5⁺ and IFN- γ ⁺CXCR5⁻CD8⁺ T cells in lymph nodes of HIV-infected patients ($n = 4$). Data are representative of two independent experiments and analysed by two-tailed paired t -test (**b**, **f**, **g**). The correlation between viral load and CXCR5⁺CD8⁺ T cell number was analysed by non-parametric Spearman correlation test (**c**). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. NS, not significant.



Extended Data Figure 10 | Diagrammatic summary of the fate of CXCR5⁺CD8⁺ T cells during chronic viral infection. During chronic viral infection, virus-specific exhausted CD8⁺ T cells differentiate into CXCR5⁺ and CXCR5⁻ subsets governed by the Id2/-E2A axis. The CXCR5⁺CD8⁺ subset migrates into B-cell follicles, where a lesser inhibitory microenvironment prevents the rapid exhaustion and loss of effector functions of these cells. By contrast, the CXCR5⁻ subset

undergoes severe exhaustion owing to the inhibitory microenvironment outside B-cell follicles. Follicular CXCR5⁺CD8⁺ T cells eventually convert into CXCR5⁻ cells, presumably driven by increased Id2 expression. The *de novo* converted CXCR5⁻CD8⁺ T cells possess better cytotoxicity, hence they are capable of clearing virus-infected cells more efficiently outside of follicles when they exit B-cell follicles.

Defining CD8⁺ T cells that provide the proliferative burst after PD-1 therapy

Se Jin Im¹, Masao Hashimoto¹, Michael Y. Gerner^{2,3}, Junghwa Lee¹, Haydn T. Kissick^{1,4}, Matheus C. Burger⁵, Qiang Shan⁶, J. Scott Hale¹, Judong Lee¹, Tahseen H. Nasti¹, Arlene H. Sharpe^{7,8}, Gordon J. Freeman⁹, Ronald N. Germain², Helder I. Nakaya⁵, Hai-Hui Xue^{6,10} & Rafi Ahmed¹

Chronic viral infections are characterized by a state of CD8⁺ T-cell dysfunction that is associated with expression of the programmed cell death 1 (PD-1) inhibitory receptor^{1–4}. A better understanding of the mechanisms that regulate CD8⁺ T-cell responses during chronic infection is required to improve immunotherapies that restore function in exhausted CD8⁺ T cells. Here we identify a population of virus-specific CD8⁺ T cells that proliferate after blockade of the PD-1 inhibitory pathway in mice chronically infected with lymphocytic choriomeningitis virus (LCMV). These LCMV-specific CD8⁺ T cells expressed the PD-1 inhibitory receptor, but also expressed several costimulatory molecules such as ICOS and CD28. This CD8⁺ T-cell subset was characterized by a unique gene signature that was related to that of CD4⁺ T follicular helper (T_{FH}) cells, CD8⁺ T cell memory precursors and haematopoietic stem cell progenitors, but that was distinct from that of CD4⁺ T_H1 cells and CD8⁺ terminal effectors. This CD8⁺ T-cell population was found only in lymphoid tissues and resided predominantly in the T-cell zones along with naive CD8⁺ T cells. These PD-1⁺CD8⁺ T cells resembled stem cells during chronic LCMV infection, undergoing self-renewal and also differentiating into the terminally exhausted CD8⁺ T cells that were present in both lymphoid and non-lymphoid tissues. The proliferative burst after PD-1 blockade came almost exclusively from this CD8⁺ T-cell subset. Notably, the transcription factor TCF1 had a cell-intrinsic and essential role in the generation of this CD8⁺ T-cell subset. These findings provide a better understanding of T-cell exhaustion and have implications in the optimization of PD-1-directed immunotherapy in chronic infections and cancer.

Functional exhaustion of antigen-specific CD8⁺ T cells has been well-documented during persistent infections^{1,2} and cancer³. A hallmark of exhausted CD8⁺ T cells is expression of various inhibitory receptors, most notably PD-1 (ref. 4). Several studies have shown that the pool of exhausted CD8⁺ T cells is phenotypically and functionally heterogeneous^{5–8}. Our goal here was to better characterize the CD8⁺ T cells that are present during chronic viral infection. A previous study shows that a subset of human CD8⁺ T cells express CXCR5 (ref. 9), a chemokine receptor, that is normally present on B cells and CD4⁺ T_{FH} cells. Another study described CXCR5⁺ CD8⁺ T cells that regulate autoimmunity in mice¹⁰. We therefore investigated whether CXCR5⁺CD8⁺ T cells were also generated during persistent viral infections. We addressed this issue using the mouse model of LCMV infection in which T-cell exhaustion was first documented¹. We found that there was a distinct population of CXCR5⁺ LCMV glycoprotein 33–41 epitope (GP33)-specific CD8⁺ T cells in the spleens

of chronically infected mice (LCMV clone 13 strain), whereas GP33-specific memory CD8⁺ T cells in mice that had cleared the infection (LCMV Armstrong strain) did not express CXCR5 (Fig. 1a). The CXCR5⁺CD8⁺ T cells in chronically infected mice also expressed the CD4⁺ T_{FH} markers ICOS and Bcl-6 and were negative for Tim-3, a marker associated with CD4⁺ T_H1 cells¹¹. In contrast, the CXCR5[−]GP33-specific CD8⁺ T cells in chronically infected mice expressed Tim-3 and were negative for ICOS and Bcl-6. Both subsets of GP33-specific CD8⁺ T cells in chronically infected mice expressed high levels of the PD-1 inhibitory receptor, with the CXCR5[−] cells showing slightly higher levels (Fig. 1a). An identical pattern of expression of these molecules was seen with CD8⁺ T cells that recognize another LCMV epitope, GP276 (Extended Data Fig. 1a). Thus, this novel population of CXCR5⁺ cells was seen with both tetramer-positive CD8⁺ T cells and these cells were detectable as early as day 8 after infection and were stably maintained in mice with high levels of viraemia (Fig. 1b, Extended Data Fig. 1b). To determine if the generation of these cells was due to antigen persistence or to the different tropism of LCMV clone 13 (ref. 12), mice were infected with either a low dose (2 × 10² plaque-forming units (PFU)) of clone 13 that is controlled within a week, or with a high dose (2 × 10⁶ PFU) that causes a persistent infection. CXCR5⁺ LCMV-specific CD8⁺ T cells were only generated in the chronically infected mice, showing that antigen persistence drives the generation of this CD8⁺ T-cell subset (Extended Data Fig. 2).

Transcriptional profiling revealed that the PD-1⁺CXCR5⁺ and PD-1⁺CXCR5[−] CD8⁺ T cells in chronically infected mice had distinct gene signatures (Extended Data Fig. 3a). Notably, the CXCR5⁺CD8⁺ T cells expressed higher levels of several costimulatory molecules (*Cd28*, *Icos*, *Tnfrsf14* (LIGHT), *Tnfrsf4* (OX-40)) and lower levels of inhibitory receptors (*Cd244* (2B4), *Havcr2* (Tim-3), *Entpd1* (CD39), *Lag3*) compared to CXCR5[−] cells (Fig. 1c, Extended Data Fig. 3b). These two CD8⁺ T-cell populations also showed differences in the expression of effector molecules, chemokines and chemokine receptors, Toll-like receptors (TLRs), transcription factors and memory markers (Fig. 1c, Extended Data Fig. 3b). CXCR5[−]CD8⁺ T cells had higher levels of several effector molecules (perforin, granzymes, etc.), but did not express IL-2 or TNF, suggesting a more terminally differentiated state (Fig. 1c, Extended Data Fig. 4), confirming and extending earlier results with PD-1⁺ Tim-3⁺ CD8⁺ T cells⁷. Interestingly, the two CD8⁺ T-cell subsets expressed different *Tlr* genes (Fig. 1c). TLRs are key molecules associated with innate immune responses, but their role on CD8⁺ T cells is not well understood¹³. *Tlr3* and *Tlr7* were selectively upregulated by CXCR5⁺CD8⁺ T cells and this was corroborated by enrichment of TLR cascade genes and interferon signalling pathways in

¹Emory Vaccine Center and Department of Microbiology and Immunology, Emory University School of Medicine, Atlanta, Georgia 30322, USA. ²Lymphocyte Biology Section, Laboratory of Systems Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892-0421, USA. ³Department of Immunology, University of Washington School of Medicine, Seattle, Washington 98109, USA. ⁴Department of Urology, Emory University School of Medicine, Atlanta, Georgia 30322, USA. ⁵School of Pharmaceutical Sciences, University of São Paulo, São Paulo 05508, Brazil. ⁶Department of Microbiology, Carver College of Medicine, University of Iowa, Iowa City, Iowa 52242, USA. ⁷Department of Microbiology and Immunology, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁸Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. ⁹Department of Medical Oncology, Dana-Farber Cancer Institute, Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. ¹⁰Interdisciplinary Immunology Graduate Program, Carver College of Medicine, University of Iowa, Iowa City, Iowa 52242, USA.

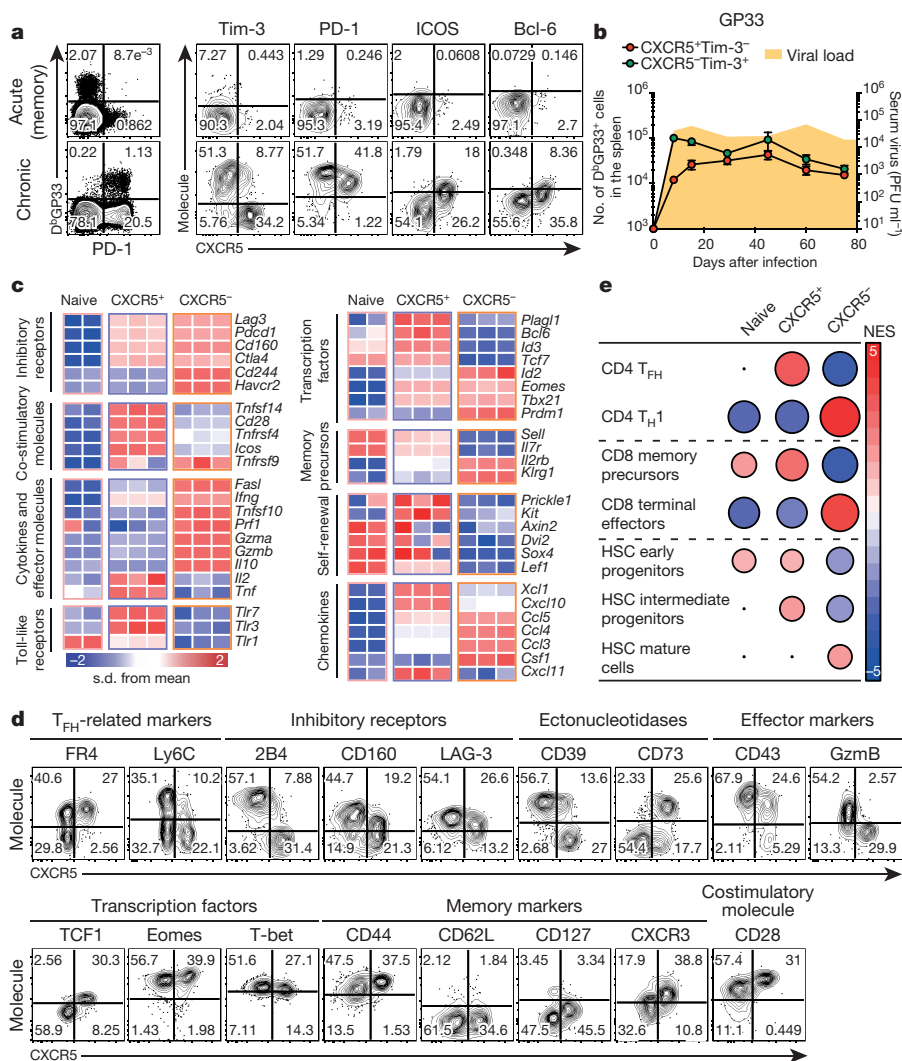


Figure 1 | Identification of a population of PD-1⁺CD8⁺ T cells during chronic LCMV infection that has a unique gene signature that resembles both CD4⁺ T_{FH} cells and CD8 memory precursor T cells. **a**, Phenotypic analysis of GP33-specific CD8⁺ T cells in the spleens of mice infected with LCMV Armstrong (acute) or clone 13 (chronic) at day 30 post-infection (p.i.). **b**, Longitudinal analysis of the numbers of GP33-specific CXCR5⁺ and CXCR5⁻ CD8⁺ T cells in the spleens of chronically infected mice ($n = 8$ from two experiments per time point). Graph shows the mean and s.e.m. **c**, Heat map illustrating the relative expression of genes in naive CD8⁺ T cells from uninfected mice and CXCR5⁺Tim-3⁻ PD-1⁺ and CXCR5⁻Tim-3⁺ PD-1⁺ CD8⁺ T cells from chronically infected mice (>45 days p.i.). **d**, Phenotypic characterization of CXCR5⁺ and CXCR5⁻ GP33-specific CD8⁺ T cells in the spleens of chronically infected mice (>45 days p.i.). **e**, GSEA for identifying specific gene signatures of the two CD8⁺ T cell subsets from chronically infected mice compared to gene signatures of different subsets of CD4⁺ and CD8⁺ T cells and haematopoietic stem cells. HSC, haematopoietic stem cell; NES, normalized enrichment score.

this subset (Extended Data Fig. 5a). Regarding transcription factors, the CXCR5⁺CD8⁺ T cells expressed *Bcl6*, *Tcf7* and *Plagl1* that are typically associated with CD4⁺ T_{FH} cells¹⁴, whereas CXCR5⁻ cells expressed *Prdm1* (Blimp-1) that is linked with CD4⁺ T_H1 cells and effector CD8⁺ T cells, highlighting the distinct transcriptional fates of these two CD8⁺ T-cell subsets. Both subsets expressed *Eomes* and *Tbx21* (T-bet) but CXCR5⁺ cells showed higher *Eomes* and lower *Tbx21* expression. The expression pattern of *Id2* and *Id3* was informative; the CXCR5⁻ cells had high *Id2* and low *Id3* pattern found in terminal effector CD8⁺ T cells that mostly die whereas CXCR5⁺ CD8⁺ T cells had low *Id2* and high *Id3*, the transcriptional profile characteristic of memory precursor CD8⁺ T cells that survive and give rise to the pool of long-lived memory cells¹⁵. The expression of memory-cell markers *Sell* (CD62L) and *Il7r* (CD127) was also consistent with the CXCR5⁺ cells being less differentiated. In addition, the CXCR5⁺ subset had enriched genes associated with mitochondrial fatty acid β -oxidation and mTOR signalling (Extended Data Fig. 5). Recent studies have highlighted the importance of fatty acid metabolism in maintenance of memory CD8⁺ T cells^{16,17}. Furthermore, CXCR5⁺CD8⁺ T cells expressed several genes in the Wnt signalling pathway that are known to be associated with self-renewal and the maintenance of haematopoietic stem cells¹⁸ (Fig. 1c). To determine if gene expression levels seen by microarray analysis correlated with protein expression, we co-stained CXCR5⁺ and CXCR5⁻ GP33-specific and GP276-specific CD8⁺ T cells with a representative set of markers and found that there was an excellent correlation between RNA levels and protein expression (Fig. 1d, Extended Data Fig. 1c). Gene set enrichment analysis (GSEA) showed

that CXCR5⁻CD8⁺ T cells were related to CD4⁺ T_H1 cells and CD8⁺ terminal effectors, whereas the CXCR5⁺ subset was similar to CD4⁺ T_{FH} cells and CD8 memory precursors (Fig. 1e, Extended Data Fig. 6). Interestingly, we also found a relationship between CXCR5⁺CD8⁺ T cells and haematopoietic stem cell progenitors (Fig. 1e). Taken together, these results suggest that LCMV-specific CXCR5⁺CD8⁺ T cells may function as memory stem cells during chronic infection.

LCMV clone 13 causes a disseminated infection that targets multiple lymphoid and non-lymphoid organs, so we next analysed the tissue distribution of the two CD8⁺ T-cell subsets in chronically infected mice. We found that LCMV-specific CXCR5⁺CD8⁺ T cells were present only in lymphoid tissues, whereas the more terminally differentiated CXCR5⁻CD8⁺ T cells were present in both lymphoid and non-lymphoid organs (Fig. 2a, Extended Data Fig. 7a, b). The blood presented a notable pattern; during the early phase (day 8) of chronic infection, both subsets were present in the blood, but later (day 30 onwards) only the CXCR5⁻CD8⁺ T cells were in circulation (Extended Data Fig. 7c, d). As both CD8⁺ T-cell subsets were present in the spleen, we determined their anatomic location within the organ. We used multiplexed confocal imaging coupled with histocytometry analysis, a technique that simultaneously permits quantitative assessment of cellular phenotype and positioning in tissues¹⁹. Owing to lack of available CXCR5 antibodies capable of *in situ* staining of fixed mouse tissues, we identified the two CD8⁺ T-cell subsets based on expression of TCF1. The CXCR5⁺Tim-3⁻ subset is TCF1⁺ whereas the CXCR5⁻Tim-3⁺ cells are TCF1⁻ (Fig. 1d). Naive CD8⁺ T cells also express TCF1 so we used the PD-1 stain to discriminate between

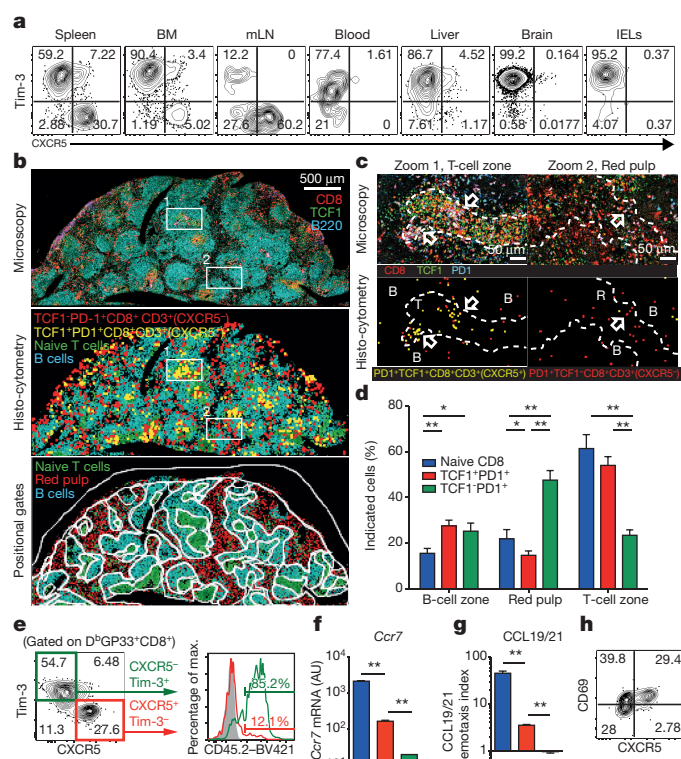


Figure 2 | CXCR5⁺PD-1⁺CD8⁺ T cells are found in lymphoid tissues of chronically infected mice and reside predominantly in T-cell zones.

a, Frequency of PD-1⁺CXCR5⁺Tim-3⁻ and PD-1⁺CXCR5⁻Tim-3⁺ LCMV-specific (GP33 + GP276) CD8⁺ T cells in the indicated tissues (45 days p.i.). **b**, Representative histology of the spleen and nuclear histo-cytometry analysis to identify the anatomic location of the two CD8⁺ T-cell subsets in chronically infected mice. **c**, Zoomed-in panels for the T-cell zone and the red pulp. T, T-cell zone; R, red pulp; B, B-cell zone. **d**, Frequency of the CD8⁺ T-cell subsets within the respective zones. Data are representative of 3 experiments ($n = 3$ per experiment). **e**, *In vivo* CD45.2 labelling of CXCR5⁺ and CXCR5⁻ GP33-specific CD8⁺ T cells in spleen, 3 min after injection ($n = 4$). **f**, Relative gene expression of *Ccr7* mRNA in sorted CD8⁺ T-cell subsets. **g**, Migration of sorted CD8⁺ T-cell subsets in response to CCL19 and CCL21. Data are combined from 2 experiments done in duplicate wells. **h**, CXCR5 and CD69 expression on GP33-specific CD8⁺ T cells. Graph shows the mean and s.e.m. Student's *t*-test, where ** $P < 0.01$; * $P < 0.05$. BM, bone marrow; mLN, mesenteric lymph nodes; IELs, intestinal epithelial lymphocytes; AU, arbitrary units.

naïve cells (PD-1⁻) and the two CD8⁺ T cell subsets from chronically infected mice (both PD-1⁺) (Extended Data Fig. 8a). Quantitative analysis showed that the CXCR5⁺CD8⁺ T-cell subset was present predominantly in the T-cell zones of the white pulp (along with naïve T cells), whereas the CXCR5⁻ subset was located mostly in the red pulp of the spleen (Fig. 2b–d). Similar results were observed when we examined the anatomic location of the two subsets in the spleen at an earlier time after infection (Extended Data Fig. 8b, c). In this context, it is worth noting that the red pulp is the major site of LCMV infection in the spleen and this is where the more terminally differentiated CD8⁺ T cells reside²⁰ (Extended Data Fig. 8d). There are also some LCMV-infected cells in the white pulp (dendritic cells and fibroblastic reticular cells), but the red-pulp macrophages are the major reservoir of infection in the spleen²¹. We next performed *in vivo* intravascular labelling²² using injection of fluorophore-conjugated anti-CD45.2 to further confirm the differential distribution of CXCR5⁺ and CXCR5⁻ CD8⁺ T cells into the T-cell zone and red pulp, respectively. We found that most of the CXCR5⁻CD8⁺ T cells were stained by the *in vivo* antibody, showing their access to the blood in the red pulp,

whereas the CXCR5⁺CD8⁺ T cells were not stained by the intravascular staining consistent with their preferential localization in the splenic white pulp (Fig. 2e). Notably, the CXCR5⁺CD8⁺ T cells were located predominantly in the T-cell zones and not in the B-cell areas. This was despite the fact that the CXCR5 molecule on these cells was functional and these CD8⁺ T cells were able to migrate in response to the chemokine CXCL13 in an *in vitro* assay (Extended Data Fig. 8e, f). However, the CXCR5⁺CD8⁺ T cells also expressed higher levels of *Ccr7* mRNA compared to their CXCR5⁻ counterparts and were able to migrate in response to the chemokines CCL19/21 that are present in the T-cell areas (Fig. 2f, g). This functional CCR7 could explain the positioning of this CXCR5⁺CD8⁺ T-cell subset in the T-cell zone²³. The CXCR5⁺CD8⁺ T cells also expressed higher levels of CD69 (Fig. 2h). This could also contribute to their retention in the T-cell areas²⁴. Finally, the CXCR5⁺ CD8⁺ T cells express very high levels of the chemokine-encoding gene *Xcl1* that promotes interactions with XCR1⁺ lymphoid dendritic cells that are predominantly located in the white pulp²⁵ (Fig. 1c).

To examine the *in vivo* dynamics of the two CD8⁺ T-cell subsets, we transferred congenically marked Cell-trace Violet (CTV)-labelled CXCR5⁺Tim-3⁻ and CXCR5⁻Tim-3⁺ cells from chronically infected mice into infection-matched recipients (Fig. 3a). As shown in Fig. 3b, the CXCR5⁻CD8⁺ T cells exhibited minimal to no division in the spleen or liver of recipient mice 21 days after transfer and these cells also retained their phenotype. In contrast, the CXCR5⁺CD8⁺ T cells not only underwent proliferation resulting in self-renewal, but also gave rise to the CXCR5⁻Tim-3⁺ subset. Consistent with this, there was a higher frequency of donor cells in the spleens of recipient mice that received CXCR5⁺CD8⁺ T cells (Fig. 3c). These results show that CXCR5⁻CD8⁺ T cells are terminally differentiated with limited proliferative potential, whereas the CXCR5⁺CD8⁺ T cells act as stem cells during chronic infection; they undergo a slow self-renewal and also give rise to the more terminally differentiated effector-like CD8⁺ T-cell subset that is present in both lymphoid and non-lymphoid tissues. We next tested the ability of CXCR5⁺ and CXCR5⁻ CD8⁺ T-cell subsets to respond to LCMV-clone-13 infection after transfer into naïve mice. These experiments were performed after transfer of low numbers (2,500) or high numbers (90,000) of donor cells (Extended Data Fig. 9a). Identical results were seen in both conditions. The transferred CXCR5⁻CD8⁺ T cells showed no expansion in the blood, spleen or liver, but there was vigorous expansion of the transferred CXCR5⁺CD8⁺ T cells in all of these tissues (Fig. 3d, Extended Data Fig. 9b–h). In addition, CXCR5⁺ cells once again gave rise to both the CXCR5⁺ and CXCR5⁻ CD8⁺ T-cell subsets, which further documents their proliferative capacity and stem-cell-like characteristics (Fig. 3e, Extended Data Fig. 9d, g).

PD-1 is a central regulator of CD8⁺ T-cell exhaustion, and blockade of this inhibitory pathway enhances T-cell immunity in chronic viral infections and cancer^{3,4}. To determine how these two CD8⁺ T-cell subsets would respond to PD-1 blockade, CXCR5⁺ and CXCR5⁻ CD8⁺ T cells were transferred into infection-matched, chronically infected mice, and groups of these mice were then treated with PD-L1-blocking antibody (Extended Data Fig. 10a). Blockade of the PD-1 inhibitory pathway had minimal effect on CXCR5⁻CD8⁺ T cells. In contrast, the CXCR5⁺CD8⁺ T cells responded to the PD-1 blockade and were present in significantly higher numbers in mice that were treated with the anti-PD-L1 antibody (Fig. 3f). PD-1 blockade substantially increased (>30-fold) the differentiation of CXCR5⁺CD8⁺ T cells into the CXCR5⁻CD8⁺ T-cell subset (Fig. 3g, Extended Data Fig. 10b). These results show that the proliferative burst seen after PD-1 blockade comes from this novel PD-1⁺CD8⁺ T-cell subset that we have identified.

The marked difference in the expression of TCF1 between the CXCR5⁺ and CXCR5⁻ CD8⁺ T-cell subsets (Fig. 1d) was of interest, as recent studies have shown that this transcription factor plays a role in the generation of CD4⁺ T_{FH} cells^{26,27} and also in the maintenance of haematopoietic stem cells in an undifferentiated state²⁸.

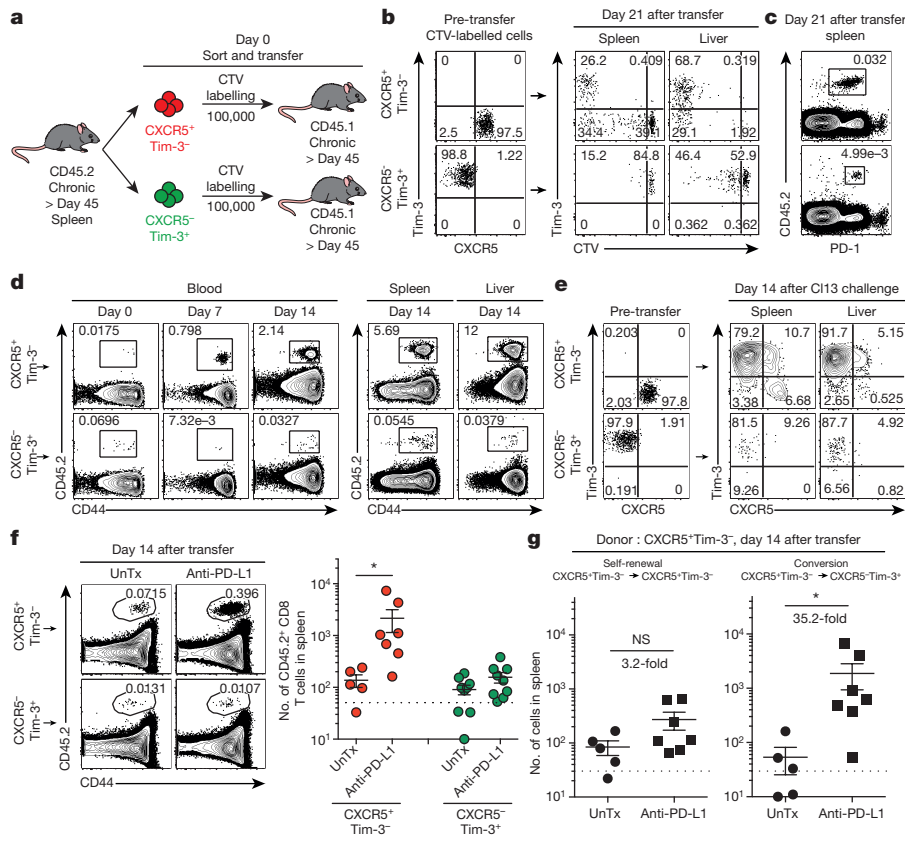


Figure 3 | CXCR5⁺CD8⁺ T cells act as stem cells during chronic LCMV infection and respond to PD-1 blockade. **a**, Experimental set-up showing transfer of congenically marked CTV-labelled PD-1⁺CXCR5⁺Tim-3⁻ and PD-1⁺CXCR5⁻Tim-3⁺ CD8⁺ T cells from chronically infected mice into infection-matched recipient mice. **b**, Proliferation and differentiation of CXCR5⁺CD8⁺ T cells in spleen and liver after adoptive transfer. **c**, FACS plots of donor CD8⁺ T cells in the spleen. Data are representative of 3 experiments (total $n = 7$ or 14 mice per group). **d**, Representative FACS plots showing expansion of donor CXCR5⁺CD8⁺ T cells after transfer into naive mice followed by LCMV clone 13 infection (see Extended Data Fig. 9a for experimental setup). **e**, Phenotypic analysis of donor CD8⁺ T-cell subsets pre-transfer and 14 days after transfer and clone 13 infection. Data are representative of 3 experiments ($n = 4$ or 6 per experiment). **f**, FACS plots and total number of donor CD8⁺ T cells recovered after adoptive transfer of CXCR5⁺ and CXCR5⁻ CD8⁺ T cells into infection-matched recipient mice followed by PD-1 blockade (see Extended Data Fig. 10a for experimental setup). **g**, Expansion and differentiation of CXCR5⁺CD8⁺ T cells after PD-1 blockade. Dashed line indicates the limit of detection. Data are combined results of 2 experiments and are mean \pm s.e.m. Student's t -test, where $*P < 0.05$.

We examined the role of TCF1 (encoded by *Tcf7*) in the generation of this CXCR5⁺CD8⁺ T-cell subset using *Tcf7*-deficient P14 transgenic CD8⁺ T cells that recognize the GP33 epitope from LCMV glycoprotein. Wild-type or *Tcf7*^{-/-} P14 cells were transferred into congenically distinct naive mice followed by LCMV clone 13 infection (Fig. 4a). The *Tcf7*^{-/-} P14 cells expanded after clone 13 infection, but exhibited a notable defect in their ability to generate CXCR5⁺Tim-3⁻ CD8⁺ T cells, whereas wild-type P14 cells gave rise to both CXCR5⁺

and CXCR5⁻ CD8⁺ T cells as expected (Fig. 4b, c, e). Endogenous GP33-specific CD8⁺ T cells in mice that received *Tcf7*^{-/-} P14 cells also differentiated normally into both CD8⁺ T-cell subsets (Fig. 4d). Taken together, these results show that TCF1 has an essential and cell intrinsic role in the differentiation of CXCR5⁺CD8⁺ T cells. The inability to generate this CD8⁺ T-cell subset was coupled to a marked loss of the *Tcf7*^{-/-} P14 cells from both lymphoid and non-lymphoid tissues (Fig. 4f–i). In summary, these data show that TCF1 is indispensable for

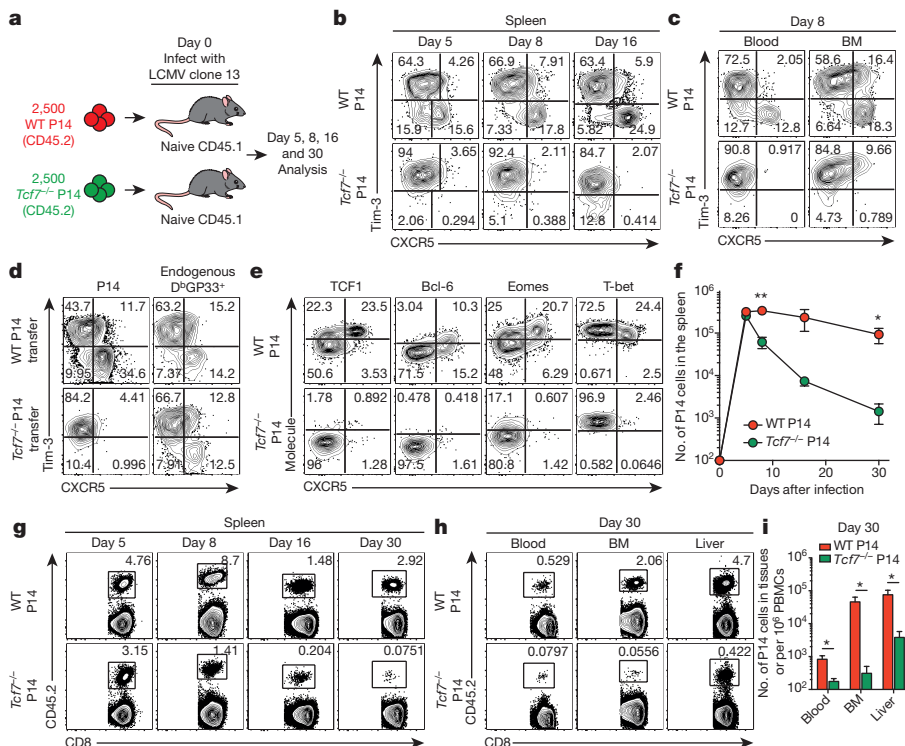


Figure 4 | TCF1 is required for the generation of CXCR5⁺CD8⁺ T cells during chronic LCMV infection. **a**, Congenically marked wild-type or *Tcf7*^{-/-} LCMV GP33-specific P14 transgenic CD8⁺ T cells were transferred into naive mice, followed by LCMV clone 13 infection. **b**, **c**, CXCR5 and Tim-3 expression on wild-type and *Tcf7*^{-/-} P14 cells in the spleen (**b**) and in tissues (**c**) at the indicated time points. **d**, CXCR5 and Tim-3 expression on wild-type or *Tcf7*^{-/-} donor P14 cells and on endogenous GP33-specific CD8⁺ T cells in the spleen 20 days p.i. **e**, Phenotypic analysis of wild-type and *Tcf7*^{-/-} P14 cells in the spleen 8 days p.i. **f**, **g**, Longitudinal analysis of the numbers (**f**) and FACS plots (**g**) of wild-type and *Tcf7*^{-/-} P14 cells in the spleen after clone 13 infection. **h**, **i**, FACS plots (**h**) and total number (**i**) of wild-type and *Tcf7*^{-/-} P14 cells in the tissues 30 days p.i. Data are representative of 2 experiments ($n = 7$ or 8 mice per time point). Graph shows the mean and s.e.m. Student's t -test, where $**P < 0.01$; $*P < 0.05$.

the generation of CXCR5⁺CD8⁺ T cells and that these stem-like cells are critical for the maintenance of virus-specific CD8⁺ T cells during chronic infection.

We have defined a PD-1⁺ virus-specific CD8⁺ T-cell population in chronically infected mice that is characterized by a unique gene signature with similarities to CD4⁺ T_{FH} cells, CD8 memory precursor cells and haematopoietic stem cell progenitors. This unique transcriptional program may represent a specific adaptation of CD8⁺ T cells to chronic antigenic stimulation. It will be of interest to determine if a similar adaptation occurs during autoimmunity and cancer. The identification of such CD8⁺ T cells in cancer will be of special relevance as our studies in chronic LCMV infection have shown that these CD8⁺ T cells selectively proliferate after PD-1 blockade. PD-1-directed immunotherapy is now one of the most promising approaches for treatment of several different types of cancers and is an approved drug for melanoma, lung cancer and bladder cancer. Our study, defining the phenotype and gene expression program of the CD8⁺ T cells that respond to PD-1 blockade, should facilitate the rational design of combination immune therapies.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 February; accepted 27 July 2016.

Published online 2 August 2016.

1. Zajac, A. J. *et al.* Viral immune evasion due to persistence of activated T cells without effector function. *J. Exp. Med.* **188**, 2205–2213 (1998).
2. Gallimore, A. *et al.* Induction and exhaustion of lymphocytic choriomeningitis virus-specific cytotoxic T lymphocytes visualized using soluble tetrameric major histocompatibility complex class I-peptide complexes. *J. Exp. Med.* **187**, 1383–1393 (1998).
3. Sharma, P. & Allison, J. P. The future of immune checkpoint therapy. *Science* **348**, 56–61 (2015).
4. Barber, D. L. *et al.* Restoring function in exhausted CD8 T cells during chronic viral infection. *Nature* **439**, 682–687 (2006).
5. Blackburn, S. D., Shin, H., Freeman, G. J. & Wherry, E. J. Selective expansion of a subset of exhausted CD8 T cells by α PD-L1 blockade. *Proc. Natl Acad. Sci. USA* **105**, 15016–15021 (2008).
6. Blackburn, S. D. *et al.* Coregulation of CD8⁺ T cell exhaustion by multiple inhibitory receptors during chronic viral infection. *Nat. Immunol.* **10**, 29–37 (2009).
7. Jin, H. T. *et al.* Cooperation of Tim-3 and PD-1 in CD8 T-cell exhaustion during chronic viral infection. *Proc. Natl Acad. Sci. USA* **107**, 14733–14738 (2010).
8. Paley, M. A. *et al.* Progenitor and terminal subsets of CD8⁺ T cells cooperate to contain chronic viral infection. *Science* **338**, 1220–1225 (2012).
9. Quigley, M. F., Gonzalez, V. D., Granath, A., Andersson, J. & Sandberg, J. K. CXCR5⁺CCR7⁺CD8 T cells are early effector memory cells that infiltrate tonsil B cell follicles. *Eur. J. Immunol.* **37**, 3352–3362 (2007).
10. Kim, H. J., Verbinen, B., Tang, X., Lu, L. & Cantor, H. Inhibition of follicular T-helper cells by CD8⁺ regulatory T cells is essential for self tolerance. *Nature* **467**, 328–332 (2010).
11. Monney, L. *et al.* Th1-specific cell surface protein Tim-3 regulates macrophage activation and severity of an autoimmune disease. *Nature* **415**, 536–541 (2002).
12. Matloubian, M., Kolhekar, S. R., Somasundaram, T. & Ahmed, R. Molecular determinants of macrophage tropism and viral persistence: importance of single amino acid changes in the polymerase and glycoprotein of lymphocytic choriomeningitis virus. *J. Virol.* **67**, 7340–7349 (1993).
13. Takeda, K., Kaisho, T. & Akira, S. Toll-like receptors. *Annu. Rev. Immunol.* **21**, 335–376 (2003).
14. Crotty, S. Follicular helper CD4 T cells (T_{FH}). *Annu. Rev. Immunol.* **29**, 621–663 (2011).
15. Yang, C. Y. *et al.* The transcriptional regulators Id2 and Id3 control the formation of distinct memory CD8⁺ T cell subsets. *Nat. Immunol.* **12**, 1221–1229 (2011).
16. Pearce, E. L. *et al.* Enhancing CD8 T-cell memory by modulating fatty acid metabolism. *Nature* **460**, 103–107 (2009).
17. Cui, G. *et al.* IL-7-induced glycerol transport and TAG synthesis promotes memory CD8⁺ T cell longevity. *Cell* **161**, 750–761 (2015).
18. Reya, T. *et al.* A role for Wnt signalling in self-renewal of haematopoietic stem cells. *Nature* **423**, 409–414 (2003).
19. Gerner, M. Y., Kastenmuller, W., Ifrim, I., Kabat, J. & Germain, R. N. Histocytometry: a method for highly multiplex quantitative tissue imaging analysis applied to dendritic cell subset microanatomy in lymph nodes. *Immunity* **37**, 364–376 (2012).
20. Jung, Y. W., Rutishauser, R. L., Joshi, N. S., Haberman, A. M. & Kaech, S. M. Differential localization of effector and memory CD8 T cell subsets in lymphoid organs during acute viral infection. *J. Immunol.* **185**, 5315–5325 (2010).
21. Mueller, S. N. *et al.* Viral targeting of fibroblastic reticular cells contributes to immunosuppression and persistence during chronic infection. *Proc. Natl Acad. Sci. USA* **104**, 15430–15435 (2007).
22. Anderson, K. G. *et al.* Intravascular staining for discrimination of vascular and tissue leukocytes. *Nat. Protocols* **9**, 209–222 (2014).
23. Haynes, N. M. *et al.* Role of CXCR5 and CCR7 in follicular T_H cell positioning and appearance of a programmed cell death gene-1^{high} germinal center-associated subpopulation. *J. Immunol.* **179**, 5099–5108 (2007).
24. Shiow, L. R. *et al.* CD69 acts downstream of interferon- α /beta to inhibit S1P1 and lymphocyte egress from lymphoid organs. *Nature* **440**, 540–544 (2006).
25. Yamazaki, C. *et al.* Critical roles of a dendritic cell subset expressing a chemokine receptor, XCR1. *J. Immunol.* **190**, 6071–6082 (2013).
26. Xu, L. *et al.* The transcription factor TCF-1 initiates the differentiation of T_{FH} cells during acute viral infection. *Nat. Immunol.* **16**, 991–999 (2015).
27. Choi, Y. S. *et al.* Lef-1 and TCF-1 orchestrate T_{FH} differentiation by regulating differentiation circuits upstream of the transcriptional repressor Bcl6. *Nat. Immunol.* **16**, 980–990 (2015).
28. Wu, J. Q. *et al.* Tcf7 is an important regulator of the switch of self-renewal and differentiation in a multipotential hematopoietic cell line. *PLoS Genet.* **8**, e1002565 (2012).

Acknowledgements This work was supported by National Institutes of Health grants R01 AI30048 (R.A.), P01 AI056299 (R.A. and A.H.S.), R01 AI112579 (H.H.X.) and R01 AI121080 (H.-H.X.) and also by the Intramural Research Program of NIAID, NIH (R.N.G. and M.Y.G.). H.T.K. is supported by funding from the Prostate Cancer Foundation and Swim Across America. H.I.N. receives a CNPq research fellowship. The authors acknowledge technical support from R. Karaffa and S. Durham for cell sorting.

Author Contributions R.A., S.J.I., and J.S.H. designed and analysed the experiments. S.J.I., M.H., Jun.L., Jud.L. and T.H.N. performed experiments. S.J.I., H.T.K., M.C.B. and H.I.N. analysed microarray data. M.Y.G. performed immunofluorescence staining and M.Y.G. and R.N.G. analysed data. Q.S., H.-H.X., A.H.S., and G.J.F. contributed critical materials. R.A. and S.J.I. wrote the manuscript, with all authors contributing to writing and providing feedback.

Author Information The microarray data are available in the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo>) under the accession number GSE84105. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.A. (rahmed@emory.edu).

METHODS

Mice, viral infections and virus titration. Six- to eight-week-old female C57BL/6 mice and CD45.1 congenic mice were purchased from Jackson Laboratory. Mice were infected with either LCMV Armstrong strain (2×10^5 PFU, intraperitoneally (i.p.)), low-dose LCMV clone 13 strain (2×10^2 PFU, intravenously (i.v.)) for acute infections, or high-dose LCMV clone 13 strain (2×10^6 PFU, i.v.) for chronic infections. Additionally, transient CD4⁺ T-cell depletion was used in the chronic LCMV infection model to induce life-long systemic infection with high levels of viraemia, which provides an optimal model to study T-cell exhaustion²⁹. Serum viral titers were determined by plaque assay on Vero E6 cells as described previously³⁰. The conditional knockout P14 female mice for *Tcf7* were Rosa26^{GFP}*Tcf7*^{fl/fl}hCD2-Cre⁺ (referred to as *Tcf7*^{-/-}) in which 80–90% of peripheral T cells had deletion of *Tcf7* with GFP expression representing Cre recombinase activity²⁷. Litter-mate Rosa26^{GFP}*Tcf7*^{fl/fl}hCD2-Cre⁻ P14 mice were used as wild-type control mice. LCMV D^bGP33-specific TCR transgenic P14 mice were fully backcrossed to C56BL/6 mice. No statistical methods were used to predetermine sample size. The number of animals for each experiment was determined based on previous experience with the model system. The investigators were not blinded to allocation during experiments and outcome assessment and the experiments were not randomized. All animal experiments were performed in accordance with Emory University Institutional Animal Care and Use Committee.

Flow cytometry. Flow cytometric analysis was performed on a FACS Canto II or LSR II (BD Biosciences). Lymphocytes were isolated from tissues including spleen, blood, liver, bone marrow, brain, gut intestinal epithelium and mesenteric lymph nodes as described previously^{30,31}. Direct *ex vivo* staining and intracellular cytokine staining were performed as described previously³⁰ with fluorochrome-conjugated antibodies (purchased from BD Bioscience, eBioscience, BioLegend, R&D, Cell Signaling Technology, Vector Laboratories and Invitrogen). To detect LCMV-specific CD8⁺ T-cell responses, tetramers were prepared as described previously³². For detection of CXCR5, a three-step staining protocol was used as described previously³³ with minor modifications. Cells were stained with tetramer and rat anti-mouse CXCR5 antibody (BD Bioscience). Samples were then incubated with 20 μ M d-biotin (Avidity) and a secondary biotin-SP-conjugated Affinipure F(Ab)₂ goat anti-rat IgG (Jackson Immunoresearch). Finally, cells were stained with streptavidin-APC (Invitrogen), streptavidin-PE or streptavidin-BV421 (BioLegend) as well as with antibodies specific to surface molecules. Note that collagenase digestion resulted in reduced staining for CXCR5. For intracellular detection of transcription factors such as Bcl-6, T-bet, Eomes and TCF1, surface-stained cells were permeabilized, fixed and stained by using the Foxp3 Permeabilization/Fixation Kit according to manufacturer's instructions (eBioscience). For intracellular detection of pS6, surface-stained cells were permeabilized, fixed and stained using Phosflow Lyse/Fix buffer (BD) and Phosflow Perm/Wash buffer I (BD). For *in vivo* antibody labelling, 30 μ g of BV421-conjugated anti-CD45.2 antibody (BioLegend) was injected i.v. into chronically infected mice. Three minutes after the injection, splenocytes were isolated and used for direct *ex vivo* staining as described previously²². FACS data were analysed with FlowJo software (TreeStar).

Cell sorting. Cell sorting was performed on a FACS Aria II (BD Biosciences). Microarray analysis, *in vitro* chemotaxis, and transfer experiments were performed on PD-1⁺CXCR5⁺Tim-3⁻ and PD-1⁺CXCR5⁺Tim-3⁺ CD8⁺ T cells sorted from chronically infected mice (>45 days p.i.) at a purity of greater than 96%. CD44^{lo}CD8⁺ T cells and B220⁺CD19⁺ B cells were isolated from uninfected mice.

RNA isolation and microarray analysis. RNA from sorted cells was purified (QIAGEN) and hybridized to Affymetrix mouse 430 2.0 arrays (Memorial Sloan Kettering Cancer Center, Genomics Core Facility). Raw data (CEL files) were normalized by RMA using Affy R package. Principal component analysis was performed using arrayQualityMetrics R package. Differential expression analysis was performed between any two subsets using limma R package (Adjusted *P* value < 0.05 and fold-change > 1.5). Gene Set Enrichment Analysis (GSEA) was run for each cell subset in pre-ranked list mode with 1,000 permutations (nominal *P*-value cutoff < 0.01). As gene sets for the GSEA analyses, we used Reactome pathways (<http://www.reactome.org/>); the MSigDB gene sets related to haematopoietic stem cells³⁴; and gene signatures associated with T_{FH} or T_H1 CD4⁺ T cells³⁵, memory precursor/terminal effector CD8⁺ T cells³⁵ and thymic innate T_{FH}-like CD4⁺ T cells³⁶. To define these signatures, we downloaded the microarray data from GEO database (GSE16697, GSE8678 and GSE64779); collapsed probe sets that matched to the same gene symbol by taking the one with highest expression across all samples; removed genes with lowest 30% mean expression; and performed differential expression analysis between the two classes using limma (adjusted *P* value < 0.01 and fold-change > 2). Enrichment scores were visualized using the corplot package in R. Enrichment scores of Reactome pathways and the genes shared by two pathways were represented as nodes and links, respectively using Cytoscape software. The microarray data are available in the Gene Expression

Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>) under the accession number GSE84105.

Confocal microscopy and histo-cytometry. To examine the localization of the CD8⁺ T cell subsets in the spleen, 20 μ m paraformaldehyde-fixed paraffin-embedded spleen sections were prepared, imaged, and analysed as previously described with minor modifications¹⁹. Briefly, images were acquired with a Lecia SP8 tiling confocal microscope (Leica Microsystems) equipped with a 40 \times 1.3 NA oil objective. Fluorophore spillover into adjacent channels was compensated using the Leica Channel Dye Separation module. Owing to high spatial resolution of the objective, deconvolution was not performed. Nuclear histo-cytometry analysis was performed by segmentation of all Jojo-1-stained nuclei (1:15,000 dilution; Invitrogen) in the imaged volume using Imaris (Bitplane) surface creation module, and by exporting the resultant statistical information into Excel (Microsoft) and then into FlowJo 10 (TreeStar). Positional gates for the T-cell zones, B-cell follicles and the red pulp were created in FlowJo using the relative densities of CD3⁺B220⁻ T cells, CD3⁺B220⁺ B cells and CD3⁺B220⁻CD44^{hi} myeloid cells, respectively. These gates were then applied to the cells of interest to assess their relative distribution across different splenic compartments.

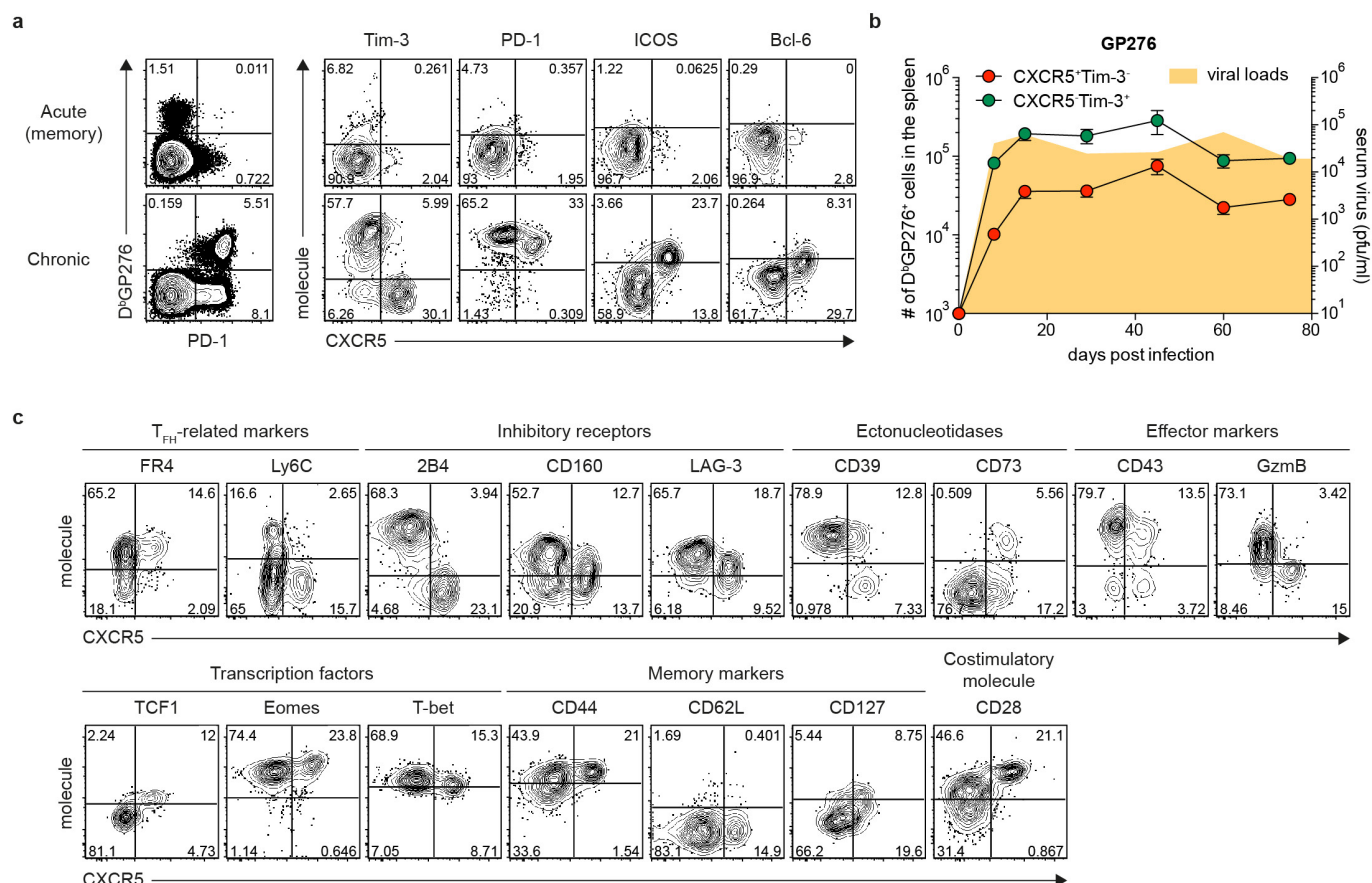
To examine which area of the spleen is infected by LCMV clone 13, spleens of chronically infected mice (>45 days p.i.) were isolated and embedded in OCT-tissue Tek and frozen immediately over liquid-nitrogen-chilled isopentane. Sections 7 μ m in thickness were cut using a micro-cryotome. They were air-dried, and then fixed in chilled acetone:methanol (1:1, v/v) at -20 °C for 10 min. The slides were permeabilized in 0.1% Triton X-100 for 30 min. The spleens were blocked with goat serum, mouse Fc block (clone 2.4G2, BD Bioscience) and avidin-biotin blocking reagent (Vectashield). The slides were then stained for LCMV antigen using pig anti-LCMV sera (1:200), BV421-rat anti-IgD (1:200, BioLegend) and biotin hamster anti-CD3 (1:100, BD Bioscience) for 1 h, followed by Alexa 555 anti-pig IgG (1:500, Invitrogen) and streptavidin-Alexa 647 (1:200, BioLegend). After washing with PBS, the slides were mounted with Prolong Gold mounting medium (Life Technologies) and cover slipped. The pictures were taken using AxioCam MRc (Zeiss) with Axionvision Rel4.7 software.

Chemotaxis. Chemotaxis assays were performed as described previously³⁷ with minor modifications. Transwells with 5- μ m pores (Corning Costar) were used. Sorted CXCR5⁺ and CXCR5⁻ CD8⁺ T cells (2.5×10^4 cells in 100 μ l) were seeded onto upper wells. The bottom wells contained either 580 μ l of PBS, a mixture of recombinant CCL19 and CCL21 (each 1 μ g ml⁻¹, R&D), or recombinant CXCL13 (3 μ g ml⁻¹, R&D). Transmigrated cells were counted by flow cytometry for 200 s at medium acquisition speed. The chemotactic index represents the ratio of cells in the lower chamber in the presence versus absence of chemokines.

Cell transfer, labelling with cell tracking dye and PD-1 blockade. For adoptive transfer experiments, 2.5×10^3 or 0.6 to 1.0×10^5 CD8⁺ T cells sorted from the spleens of chronically infected mice (>45 days p.i.) were transferred i.v. into infection-matched, or naive mice. To track the proliferation of lymphocytes, sorted CD8⁺ T cells were labelled with Cell-trace Violet (Invitrogen), according to the manufacturer's protocol. PD-1 blockade was performed as described previously⁴ after the transfer of two CD8⁺ T-cell subsets into infection-matched mice. For P14 experiments, splenocytes containing 2.5×10^3 wild-type or *Tcf7*^{-/-} P14 cells (CD45.2⁺) were transferred i.v. into naive CD45.1 recipient mice.

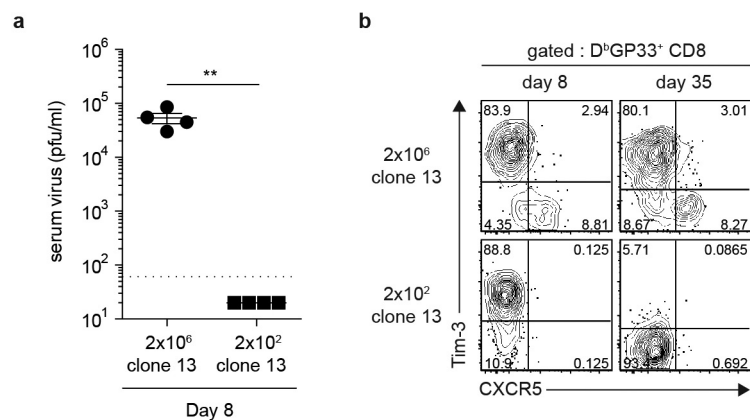
Statistical analysis. All experiments were analysed using Prism 6 (GraphPad Software). Statistical differences were assessed using a two-tailed unpaired or paired Student's *t*-test. *P* values of < 0.05 and < 0.01 indicated the significant difference between relevant groups.

- Matloubian, M., Concepcion, R. J. & Ahmed, R. CD4⁺ T cells are required to sustain CD8⁺ cytotoxic T-cell responses during chronic viral infection. *J. Virol.* **68**, 8056–8063 (1994).
- Wherry, E. J., Blattman, J. N., Murali-Krishna, K., van der Most, R. & Ahmed, R. Viral persistence alters CD8 T-cell immunodominance and tissue distribution and results in distinct stages of functional impairment. *J. Virol.* **77**, 4911–4927 (2003).
- Masopust, D. et al. Dynamic T cell migration program provides resident memory within intestinal epithelium. *J. Exp. Med.* **207**, 553–564 (2010).
- Murali-Krishna, K. et al. Counting antigen-specific CD8 T cells: a reevaluation of bystander activation during viral infection. *Immunity* **8**, 177–187 (1998).
- Johnston, R. J. et al. Bcl6 and Blimp-1 are reciprocal and antagonistic regulators of T follicular helper cell differentiation. *Science* **325**, 1006–1010 (2009).
- Ivanova, N. B. et al. A stem cell molecular signature. *Science* **298**, 601–604 (2002).
- Joshi, N. S. et al. Inflammation directs memory precursor and short-lived effector CD8⁺ T cell fates via the graded expression of T-bet transcription factor. *Immunity* **27**, 281–295 (2007).
- Miyazaki, M. et al. The E-I δ protein axis modulates the activities of the PI3K-AKT-mTORC1-Hif1 α and c-myc/p19Arf pathways to suppress innate variant T_{HH} cell development, thymocyte expansion, and lymphomagenesis. *Genes Dev.* **29**, 409–425 (2015).
- Allen, C. D. et al. Germinal center dark and light zone organization is mediated by CXCR4 and CXCR5. *Nat. Immunol.* **5**, 943–952 (2004).



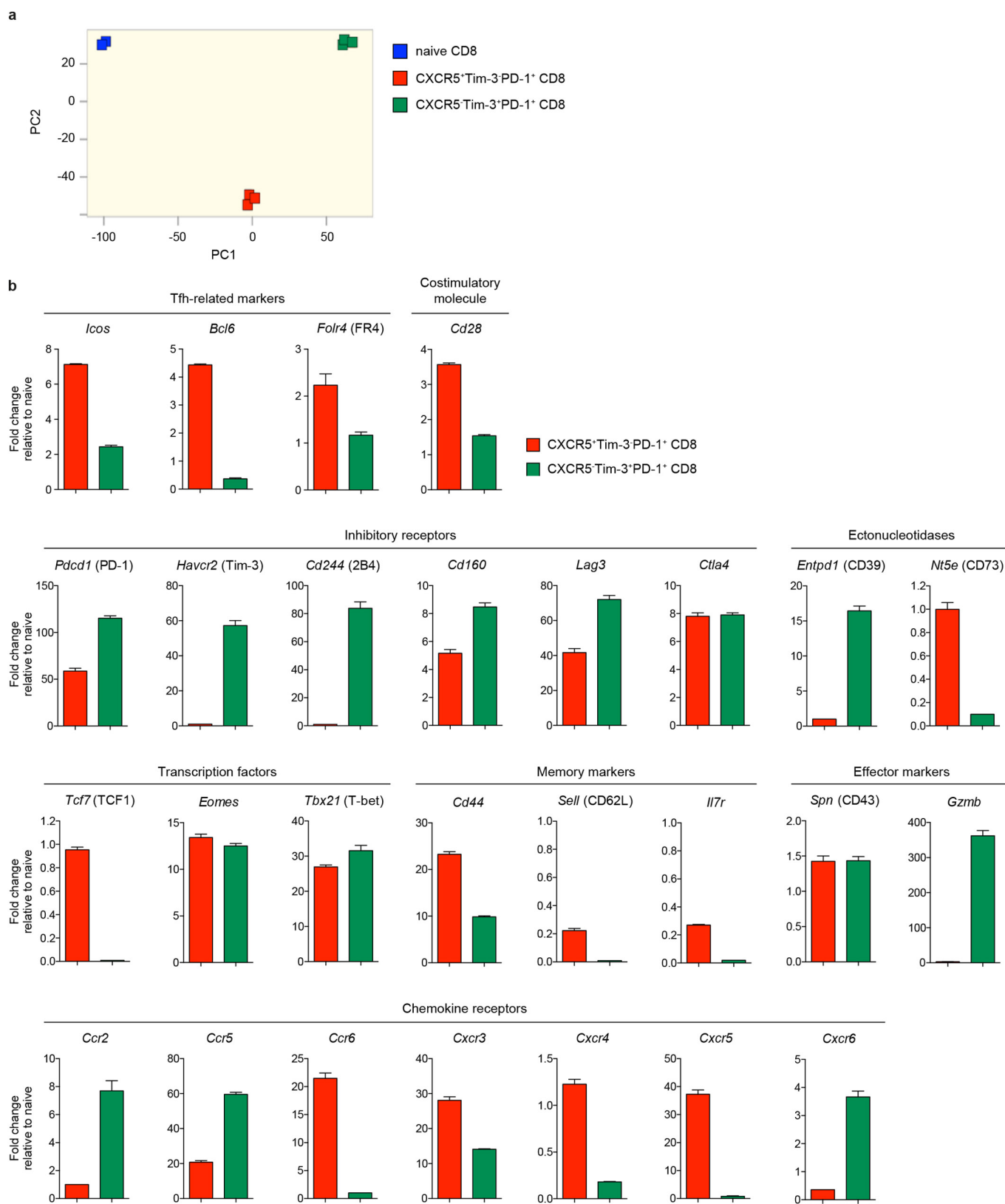
Extended Data Figure 1 | LCMV GP276-specific CD8⁺ T cells also consist of CXCR5⁺ and CXCR5⁻ CD8⁺ T-cell subsets during chronic infection. **a**, Phenotypic analysis of GP276-specific CD8⁺ T cells in the spleens of immune mice that had cleared an acute LCMV Armstrong infection or mice that were chronically infected with LCMV clone 13 (day 30 after infection (p.i.)). FACS plots showing CXCR5 expression in combination with the indicated markers are gated on GP276-tetramer⁺

CD8⁺ T cells. **b**, Longitudinal analysis of the numbers of GP276-specific CXCR5⁺Tim-3⁻ and CXCR5⁺Tim-3⁺ CD8⁺ T cells in the spleen at the indicated time after infection. LCMV titers in the serum are shown as the shaded yellow area. Graph shows the mean \pm s.e.m. Data are the average of 8 mice from two experiments per time point (total $n = 48$). **c**, Phenotypic characterization of CXCR5⁺ and CXCR5⁻ GP276-specific CD8⁺ T cells in the spleen of chronically infected mice (>45 days p.i.).



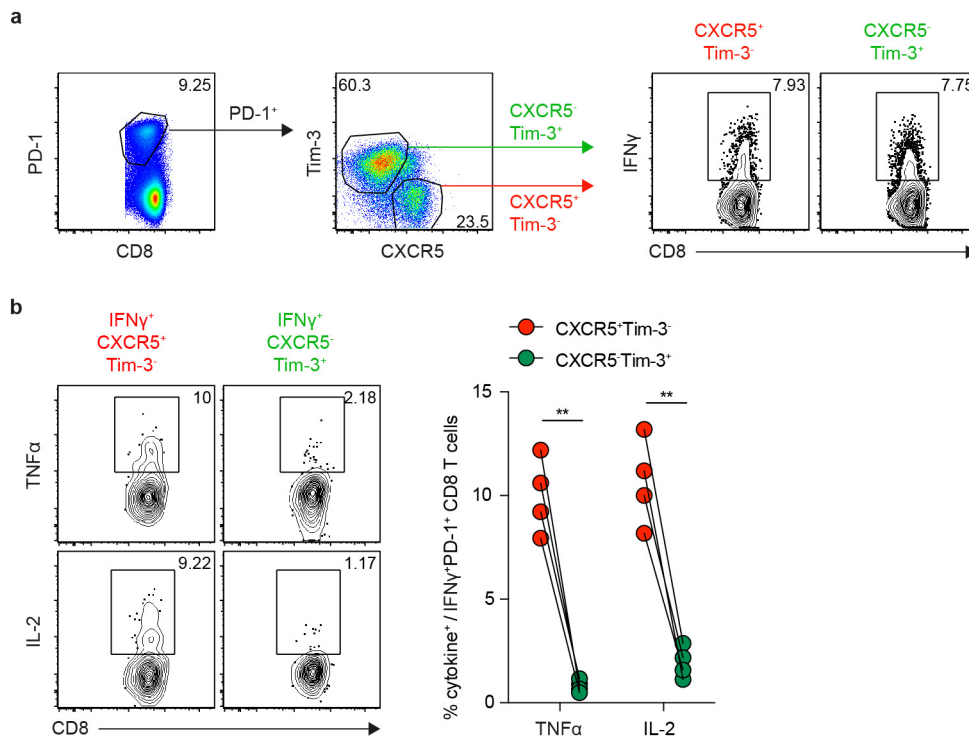
Extended Data Figure 2 | Low-dose challenge with LCMV clone 13 results in acute infection and does not generate CXCR5⁺CD8⁺ T cells. Mice were infected with either low-dose (2×10^2 PFU) or high-dose (2×10^6 PFU) of LCMV clone 13, and the generation of CXCR5⁺CD8⁺ T cells was examined at day 8 and day 35 p.i. **a**, Serum virus titers at

day 8 after infection. **b**, Representative flow plots of CXCR5 and Tim-3 expression on GP33-specific CD8⁺ T cells in the spleen at day 8 and 35 p.i. Data are obtained from a total of 16 mice with 4 mice per group at each time point. Student's *t*-test, where $**P < 0.01$.



Extended Data Figure 3 | Distinct transcriptional profiles of CXCR5⁺ and CXCR5⁻ CD8⁺ T cells from spleens of mice chronically infected with LCMV. a, Principal component analysis of naive (CD44^{lo}) CD8⁺ T cells isolated from uninfected mice and CXCR5⁺Tim-3⁺PD-1⁺ and CXCR5⁻Tim-3⁺PD-1⁺ CD8⁺ T cells isolated from chronically infected

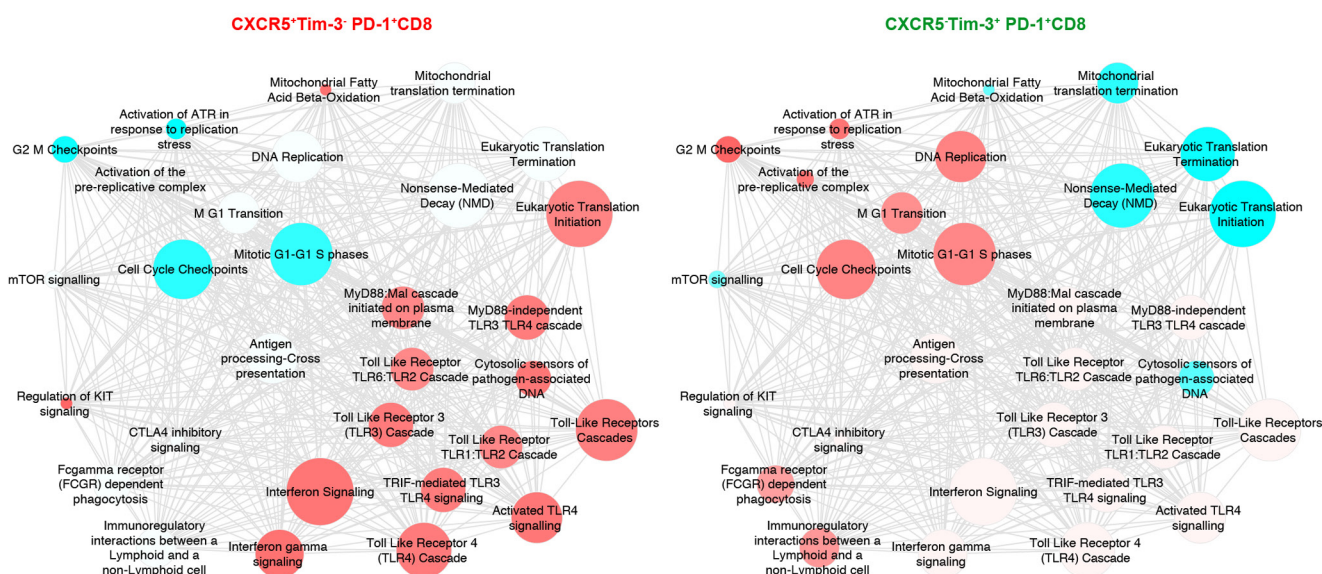
mice (>45 days p.i.). Each square represents an individual biological replicate. **b**, Relative expression of selected genes as determined by Affymetrix microarray analysis. Data are shown as fold change relative to naive (CD44^{lo}) CD8⁺ T cells. Graphs show the mean \pm s.e.m.



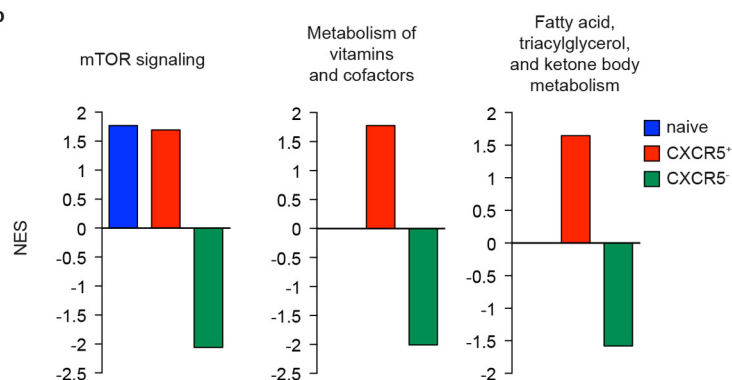
Extended Data Figure 4 | LCMV-specific CXCR5⁺CD8⁺ T cells are more polyfunctional than CXCR5⁻CD8⁺ T cells. Splenocytes from chronically infected mice (>45 days p.i.) were stimulated with GP33-41 peptide for 5 h followed by phenotypic marker staining and intracellular staining. **a**, Gating strategy for IFN-γ⁺PD-1⁺CXCR5⁺Tim-3⁻ and IFN-γ

⁺PD-1⁺CXCR5⁻Tim-3⁺CD8⁺ T cells. **b**, Representative FACS plots and graph showing the frequency of TNFα- and IL-2-producing cells among IFN-γ⁺CD8⁺ T cells. Data are representative of 2 experiments ($n = 4$ per experiment). Student's paired t -test, where $**P < 0.01$.

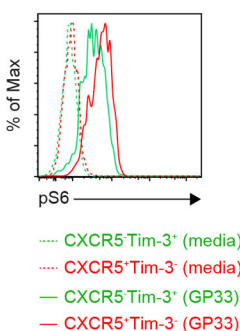
a



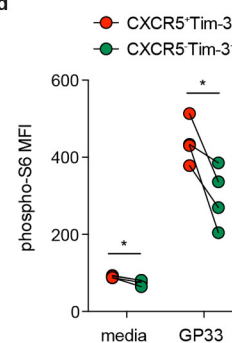
b



c

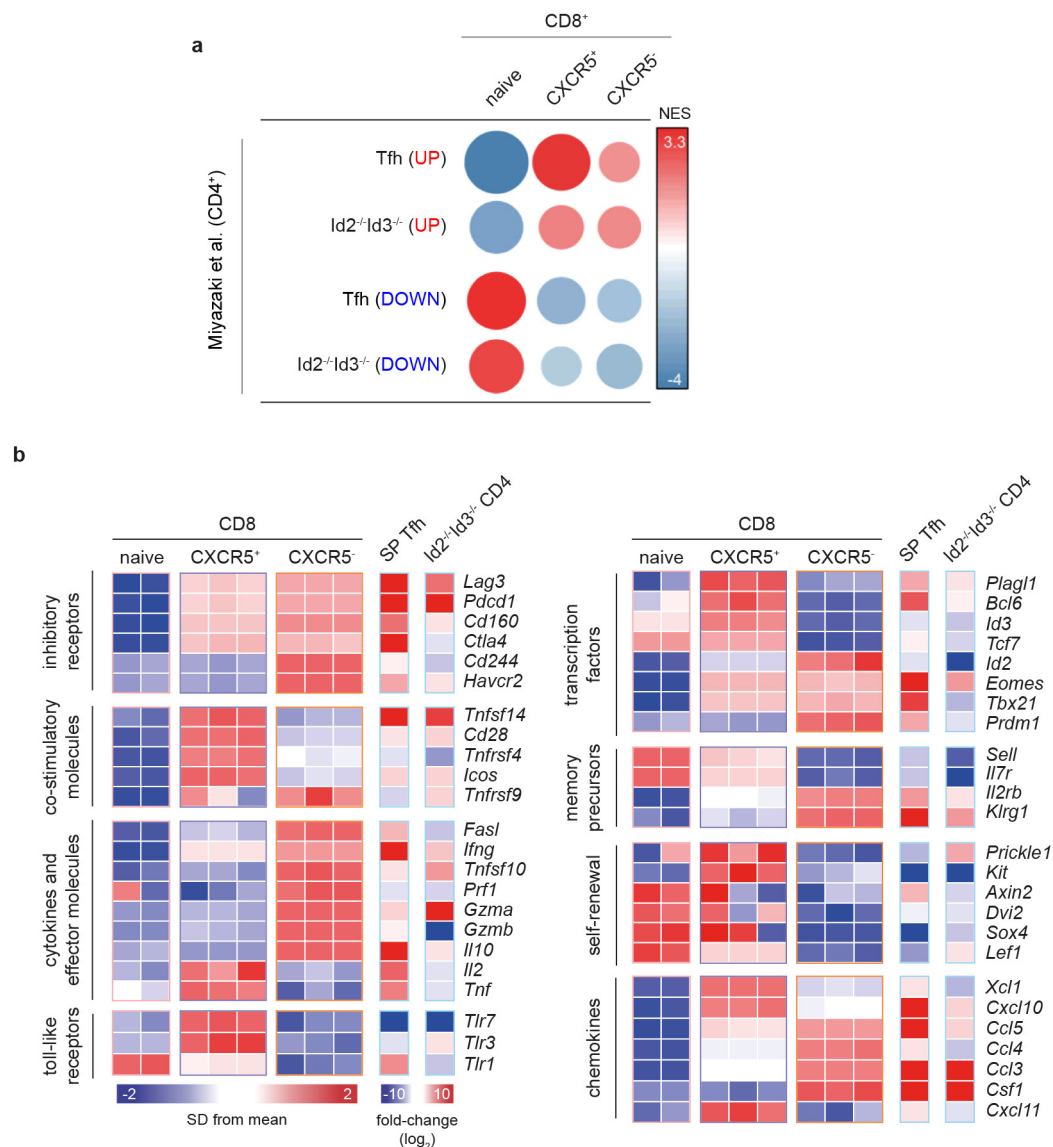


d



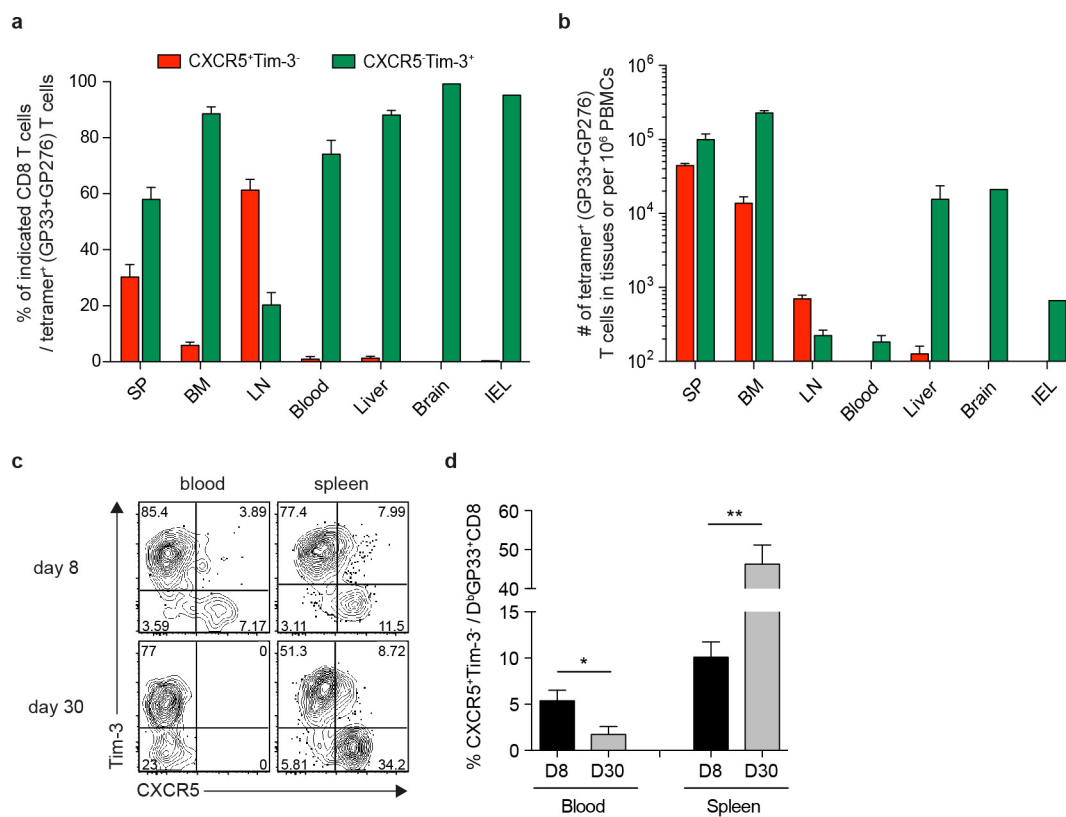
Extended Data Figure 5 | Analysis of reactome pathways, mTOR signalling and fatty acid metabolism in CXCR5⁺ and CXCR5⁻ CD8 T cells from LCMV chronically infected mice. **a**, Reactome pathways in CXCR5⁺ Tim-3⁻ PD-1⁺ and CXCR5⁻ Tim-3⁺ PD-1⁺ CD8 T cells isolated from the spleens of mice chronically infected with LCMV (>45 days p.i.). GSEA (nominal $P < 0.01$; 1,000 permutations) was used to identify positive (red, maximum normalized enrichment score (NES) = 3.2) or negative (blue, min NES = -3.7) enrichment of Reactome pathways (<http://www.reactome.org/>) in CXCR5⁺Tim-3⁻PD-1⁺ and CXCR5⁻Tim-3⁺PD-1⁺ CD8 T cells using meta-analysis. The size of the circles (nodes) represents the number of genes on each pathway. The links

between circles (edges) represent the number of genes shared by two given pathways. The networks were generated using Cytoscape. **b**, GSEA on mTOR signalling and fatty acid metabolism. Bars represent pathways with nominal P value < 0.01 . **c**, **d**, Splenocytes from chronically infected mice (>45 days p.i.) were stimulated with medium or GP33-41 peptide for 1 h followed by phenotypic marker staining and phosphorylated S6 ribosomal protein (pS6) staining. Flow cytometry analysis (**c**) and MFI (**d**) of pS6 expression in CXCR5⁺ and CXCR5⁻ CD8⁺ T-cell subsets after *ex vivo* stimulation. Data are representative of 2 experiments ($n = 4$ per experiment). Student's paired t -test, where $*P < 0.05$.



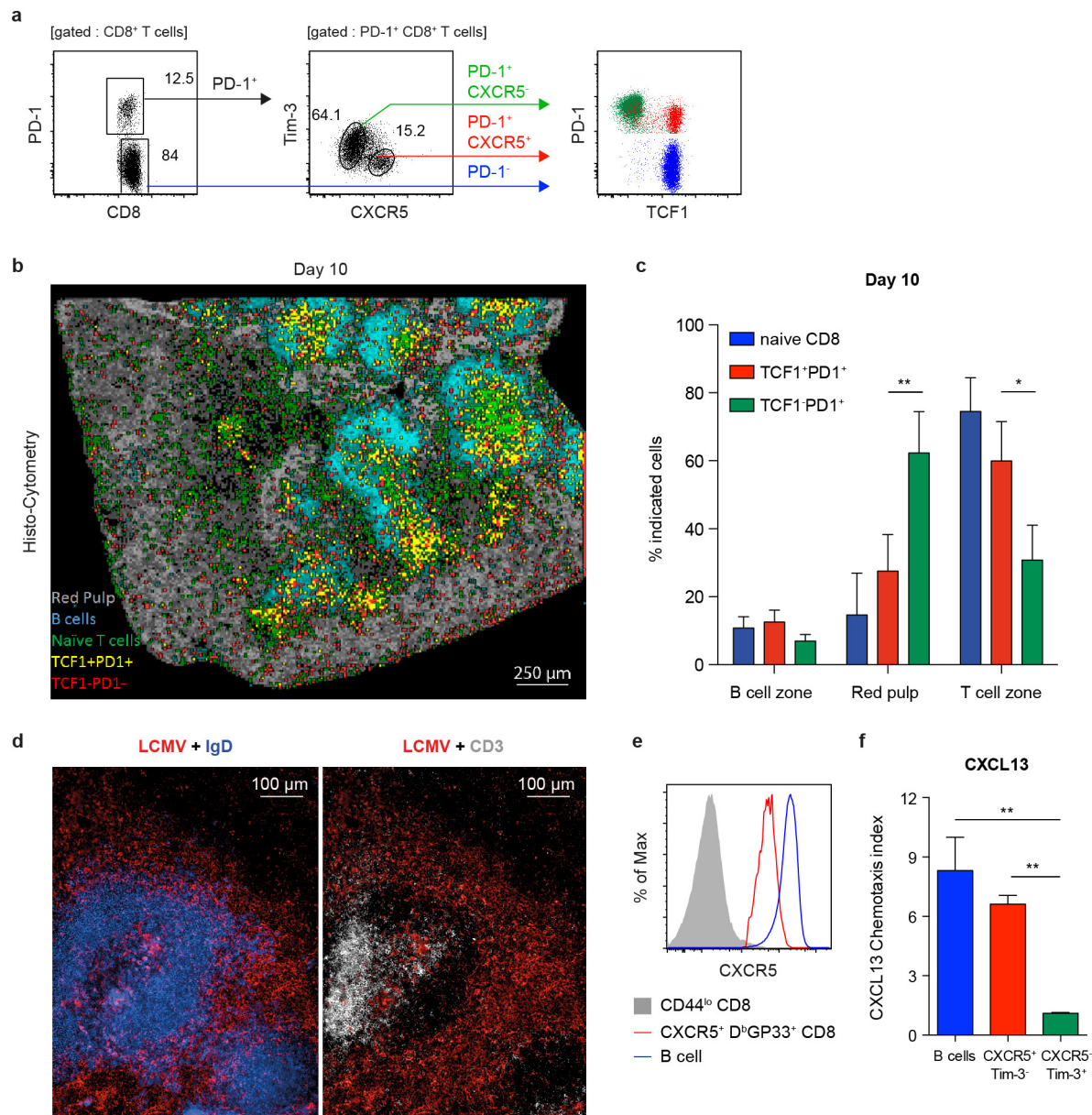
Extended Data Figure 6 | Comparison of gene signatures of CXCR5⁺ and CXCR5⁻ CD8⁺ T-cell subsets from chronically infected mice with *Id2*^{-/-}*Id3*^{-/-} innate T_{FH}-like CD4⁺ T cells. **a**, GSEA was performed using genes pre-ranked by the mean Z-score values of each CD8 subset (naive, CXCR5⁺ or CXCR5⁻) calculated across all samples. Splenic CD4⁺ T_{FH} gene signatures from wild-type mice and thymic innate variant T_{FH} gene signatures from *Id2*^{-/-}*Id3*^{-/-} mice (GSE64779) (ref. 36) were used as gene sets in our GSEA. Genes were considered up- or downregulated in cell subsets compared to control (sorted CD4⁺TCRβ⁺CD8⁻ cells) if there was a fold-change >2 and *P* < 0.05 (ref. 36). **b**, Heat map illustrating the relative expression of the indicated genes of *Id2*^{-/-}*Id3*^{-/-} CD4⁺ T cells defined in ref. 36 compared to those of naive, CXCR5⁺ and CXCR5⁻ CD8⁺ T cells. GSEA analysis revealed some interesting similarities and differences between the CXCR5⁺ CD8⁺ T cells from chronically infected mice and the *Id2*^{-/-}*Id3*^{-/-} T_{FH}-like CD4⁺

T cells. These two cell populations are distinct, but share certain biological properties such as increased self-renewal activity. For example, some of the interesting inhibitory and costimulatory molecules such as *Pdccl1* (PD-1), *Tnfrsf14* (LIGHT), *Cd28*, and *Icos* were commonly upregulated in both CXCR5⁺ CD8⁺ and innate T_{FH}-like CD4⁺ T cells, whereas molecules like *Cd244* (2B4), *Prf1*, *Fasl* and *Gzmb* were downregulated in both cell types. However, there were also many differences, perhaps the most notable being the low expression of *Tcf7* (TCF1) in the innate CD4⁺ T cells compared to the high expression of *Tcf7* in the CXCR5⁺CD8⁺ T cells and the critical role of this transcription factor in the generation of these cells. Notably, the CD4⁺ T-cell population defined in ref. 36 is genetically deficient in both *Id2* and *Id3*, whereas the CXCR5⁺CD8⁺ T cells express high levels of *Id3*. Thus, many aspects of the transcriptional program of these two cell types will be distinct.



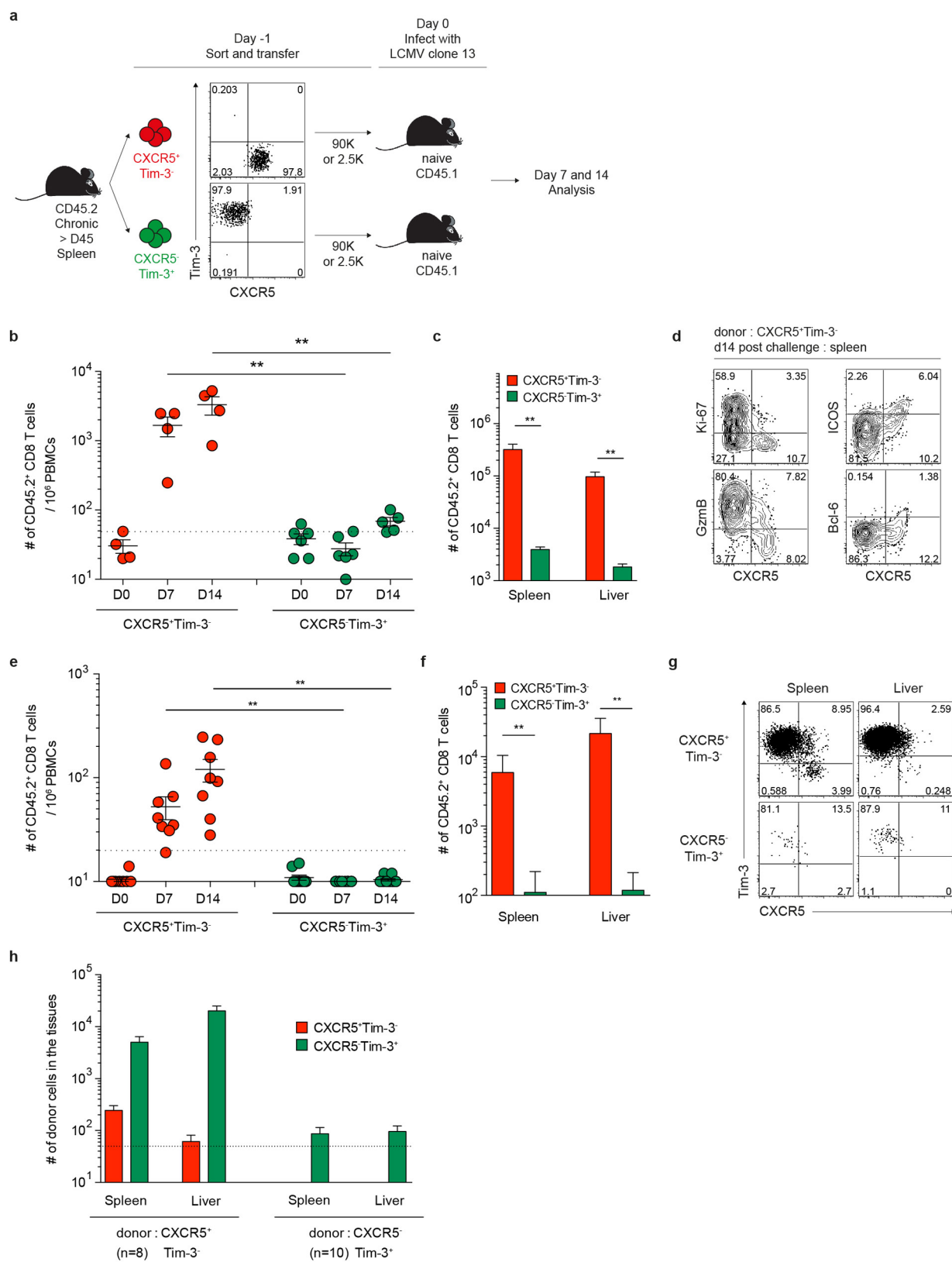
Extended Data Figure 7 | Tissue distribution of LCMV-specific CXCR5⁺ and CXCR5⁻ CD8⁺ T cells in chronically infected mice. **a, b**, Frequency (**a**) and numbers (**b**) of LCMV-specific CXCR5⁺ and CXCR5⁻ CD8⁺ T cells in the indicated tissues at >45 days p.i. in chronically infected mice (SP, spleen; BM, bone marrow; LN, mesenteric lymph nodes; IEL, intestinal epithelial lymphocytes). **c, d**, Representative

FACS plots (**c**) and graph (**d**) showing the frequency of CXCR5⁺Tim-3⁻ GP33-specific CD8⁺ T cells in the blood and spleen at day 8 and day 30 p.i. FACS plots are gated on D^bGP33 tetramer⁺ CD8⁺ T cells. Data are obtained from 4 or 8 mice. Graphs show the mean \pm s.e.m. Student's *t*-test, where ***P* < 0.01; **P* < 0.05.



Extended Data Figure 8 | Anatomic localization of CXCR5⁺ and CXCR5⁻ CD8⁺ T cells in the spleen and *in vitro* migration of the two CD8⁺ T-cell subsets to CXCL13. **a**, Analysis of TCF-1 and PD-1 expression in CD8⁺ T cell subsets in the spleen of chronically infected mice (>45 days p.i.). Left and middle FACS plots are gated on total CD8⁺ T cells and PD-1⁺ CD8⁺ T cells, respectively. Right FACS plot displays the overlay of TCF-1 and PD-1 expression in gated PD-1⁻, PD-1⁺CXCR5⁺Tim-3⁻ and PD-1⁺CXCR5⁻Tim-3⁺ CD8⁺ T cells. **b**, **c**, Nuclear histocytometry analysis of the spleen (**b**) and frequency (**c**) of the CD8⁺ T-cell subsets within the respective zones of the spleen 10 days after LCMV clone 13 infection ($n = 3$). **d**, Representative

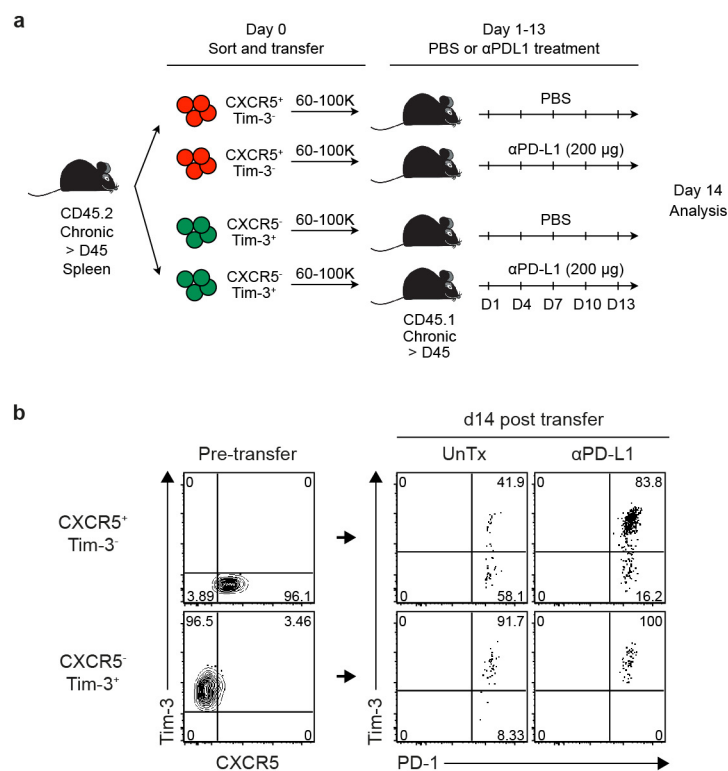
immunofluorescence staining of LCMV clone 13 infection in the spleen (>45 days p.i.). The spleen was stained for LCMV antigen (red), IgD (blue) and CD3 (white) and examined via microscopy ($n = 4$). Note, LCMV infection is mostly in the red pulp. **e**, CXCR5 expression on splenic naive (CD44^{lo}) CD8⁺ T cells, B cells and CXCR5⁺ GP33-specific CD8⁺ T cells from chronically infected mice (>45 days p.i.). **f**, Relative migration of sorted B cells and CXCR5⁺Tim-3⁻PD-1⁺ and CXCR5⁻Tim-3⁺PD-1⁺ CD8⁺ T-cell subsets in response to CXCL13, which is a ligand for CXCR5. Data are combined from two experiments performed in duplicate wells for each sample. Graphs show the mean and s.e.m. Student's *t*-test, where ** $P < 0.01$; * $P < 0.05$.



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | CXCR5⁺CD8⁺ T cells selectively undergo proliferation after LCMV clone 13 challenge. **a**, Sorted PD-1⁺CXCR5⁺Tim-3⁻ and PD-1⁺CXCR5⁻Tim-3⁺ CD8⁺ T cells isolated from CD45.2⁺ chronically infected mice (>45 days p.i.) were adoptively transferred into naive CD45.1⁺ recipient mice, followed by LCMV-clone-13 challenge. Two sets of adoptive-transfer experiments were performed; one using a low dose of donor cells (2,500 cells) and another with a large dose of donor cells (90,000 cells). The data shown in Fig. 3d, e are from the high-dose transfer experiment. Data in **b–d** are from the high-dose transfer; and in **e–h** are from the low-dose transfer. **b**, Expansion of CXCR5⁺CD8⁺ T cells in the blood after LCMV clone 13 infection. **c**, Number of cells in the spleen and liver 14 days after infection. **d**, Phenotypic analysis of transferred donor CXCR5⁺Tim-3⁻ cells in the

spleen 14 days after challenge, showing proliferation and differentiation of this subset. Data are representative of 3 experiments ($n = 4$ or 6 mice per experiment). **e**, Expansion of CXCR5⁺CD8⁺ T cells (low-dose transfer) in the blood after infection with LCMV clone 13. **f**, Number of cells in the spleen and liver 14 days after challenge. **g**, Phenotypic analysis of transferred donor CXCR5⁺Tim-3⁻ and CXCR5⁻Tim-3⁺ cells in the spleen and liver at 14 days after infection showing differentiation of CXCR5⁺CD8⁺ T cells. **h**, Number of CXCR5⁺Tim-3⁻ and CXCR5⁻Tim-3⁺ CD8⁺ T cells derived from donor CXCR5⁺ or CXCR5⁻ CD8⁺ T cells in the spleen and liver after clone 13 challenge. Dashed line indicates the limit of detection. Data are combined from two experiments ($n = 8$ or 10 per group, total $n = 18$). Graph shows the mean and s.e.m. Student's *t*-test, where $**P < 0.01$.



Extended Data Figure 10 | Enhanced conversion of CXCR5⁺Tim-3⁻CD8⁺ T cells to Tim-3⁺CD8⁺ T cells after PD-1 blockade. **a**, Sorted PD-1⁺CXCR5⁺Tim-3⁻ and PD-1⁺CXCR5⁻Tim-3⁺ CD8⁺ T cells isolated from CD45.2⁺ chronically infected mice (>45 days p.i.) were adoptively transferred into infection-matched CD45.1⁺ recipient mice, followed

by treatment with anti-PD-L1 antibody. **b**, Phenotypic analysis of sorted donor CD8⁺ T-cell subsets before transfer and 14 days after the transfer followed by PD-1 blockade. Data are representative of 2 experiments (total $n = 5, 7$, or 9 mice per group).

A PGC1 α -mediated transcriptional axis suppresses melanoma metastasis

Chi Luo^{1*}, Ji-Hong Lim^{1*†}, Yoonjin Lee^{1,2}, Scott R. Granter³, Ajith Thomas¹, Francisca Vazquez^{1,4,5}, Hans R. Widlund⁶ & Pere Puigserver¹

Melanoma is the deadliest form of commonly encountered skin cancer because of its rapid progression towards metastasis^{1,2}. Although metabolic reprogramming is tightly associated with tumour progression, the effect of metabolic regulatory circuits on metastatic processes is poorly understood. PGC1 α is a transcriptional coactivator that promotes mitochondrial biogenesis, protects against oxidative stress³ and reprograms melanoma metabolism to influence drug sensitivity and survival^{4,5}. Here, we provide data indicating that PGC1 α suppresses melanoma metastasis, acting through a pathway distinct from that of its bioenergetic functions. Elevated PGC1 α expression inversely correlates with vertical growth in human melanoma specimens. PGC1 α silencing makes poorly metastatic melanoma cells highly invasive and, conversely, PGC1 α reconstitution suppresses metastasis. Within populations of melanoma cells, there is a marked heterogeneity in PGC1 α levels, which predicts their inherent high or low metastatic capacity. Mechanistically, PGC1 α directly increases transcription of ID2, which in turn binds to and inactivates the transcription factor TCF4. Inactive TCF4 causes downregulation of metastasis-related genes, including integrins that are known to influence invasion and metastasis^{6–8}. Inhibition of BRAF^{V600E} using vemurafenib⁹, independently of its cytostatic effects, suppresses metastasis by acting on the PGC1 α –ID2–TCF4–integrin axis. Together, our findings reveal that PGC1 α maintains mitochondrial energetic metabolism and suppresses metastasis through direct regulation of parallel acting transcriptional programs. Consequently, components of these circuits define new therapeutic opportunities that may help to curb melanoma metastasis.

Whereas the landscape of genetic alterations and multiple driver mutations have been discovered in melanoma^{10,11}, less is understood about the genes that drive metastasis². Nevertheless, it is thought that efficient metastasis requires the malignant cell to balance proliferation with invasion and migration^{12,13}. Elevated expression of the metabolic integrator and transcriptional coactivator peroxisome proliferator-activated receptor- γ coactivator-1 α (PGC1 α encoded by *PPARGC1A*) defines a subset of melanomas in which it promotes mitochondrial metabolism, protects against oxidative stress and enhances survival^{4,5}. Although high PGC1 α expression is associated with worse prognosis in metastatic melanomas^{4,5}, reduced levels coincide with invasive or vertical growth in primary specimens (Fig. 1a). Therefore, we investigated the effects of PGC1 α on invasion and metastasis. Gene set enrichment analysis (GSEA) of expression data (GSE36879)⁴ upon PGC1 α knockdown in the poorly metastatic melanoma cell line A375P revealed coordinated upregulation of genes implicated in metastasis, including genes that control focal adhesion or extracellular matrix (ECM) interactions, integrins and components of the transforming

growth factor- β (TGF β) and Wnt signalling pathways^{14–17} (Extended Data Fig. 1a, b). In addition, PGC1 α expression showed an inverse correlation with gene sets involved in melanoma metastasis (Extended Data Fig. 1c, d). Upregulation of pro-metastatic genes following PGC1 α suppression was confirmed by qPCR in PGC1 α -positive melanoma cell lines (Fig. 1b and Extended Data Fig. 2a–d). Conversely, the increase in integrin transcripts was reversed upon ectopic PGC1 α expression (Extended Data Fig. 2e, f). Targeting PGC1 α using the CRISPR/Cas9 system led to similar gene expression changes (Extended Data Fig. 2g).

Consistent with lower PGC1 α expression during vertical growth and acquisition of the metastatic phenotype, the PGC1 α -suppressed invasive and metastatic gene signature was associated with worse survival in patients with primary melanoma (Fig. 1c). Changes in integrin expression upon PGC1 α depletion were accompanied by activation of the downstream focal adhesion kinase (FAK)¹⁸ (Fig. 1d) and increases in migration and invasion (Fig. 1e). FAK inhibition blocked the enhanced migration induced by PGC1 α depletion (Extended Data Fig. 2h–j). Remarkably, silencing of PGC1 α by either short hairpin RNA (shRNA) (Fig. 1f and Extended Data Fig. 3a, b) or CRISPR/Cas9 (Fig. 1g) converted these low-invasive, PGC1 α -positive cells into highly metastatic entities as assessed by tail-vein injection experiments. To fully recapitulate the metastatic process *in vivo*, we used the human melanoma cell line MeWo in an orthotopic metastasis model¹⁹ and found that, again, PGC1 α suppression caused the subcutaneously implanted tumours to generate widespread disease (Fig. 1h). Conversely, reconstitution of PGC1 α in the PGC1 α -negative cell lines A375 and A2058 decreased integrin expression (Fig. 1i and Extended Data Fig. 3c, d) and compromised their invasiveness *in vitro* and *in vivo* (Fig. 1j, k). Together, these results indicate that PGC1 α inhibits a pro-metastatic program in melanoma cells resulting in the suppression of invasion and metastasis.

Melanomas are highly heterogeneous and might switch between a proliferative and an invasive or metastatic phenotype¹³. Using MitoTracker (to label mitochondria) and FACS analysis, we found that melanoma cell lines with heightened PGC1 α expression displayed heterogeneous mitochondrial mass (Fig. 2a), which was dynamically regulated (Extended Data Fig. 4a), and therefore could implicate alternate mitochondrial biogenesis and PGC1 α function during phenotype switching. The sorted mitochondria-high (mito/PGC1 α -high) population showed significantly higher expression of PGC1 α and mitochondrial components, as well as lower expression of integrins, compared to the mitochondria-low (mito/PGC1 α -low) population (Fig. 2a, b). The PGC1 α -low population showed enhanced migration *in vitro* and metastasis *in vivo* (Fig. 2c, d). Compared to the non-migrating population, melanoma cells that had migrated through the transwell membrane expressed lower amounts of PGC1 α and higher amounts of pro-metastatic transcripts (Fig. 2e and Extended Data Fig. 4b). To strengthen the link

¹Department of Cancer Biology, Dana-Farber Cancer Institute and Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ³Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁴Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. ⁵Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁶Department of Dermatology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. [†]Present address: Department of Biomedical Chemistry, College of Biomedical and Health Science, Konkuk University, Chungbuk, Chungju, Chungcheongbuk-do 380-701, South Korea.

*These authors contributed equally to this work.

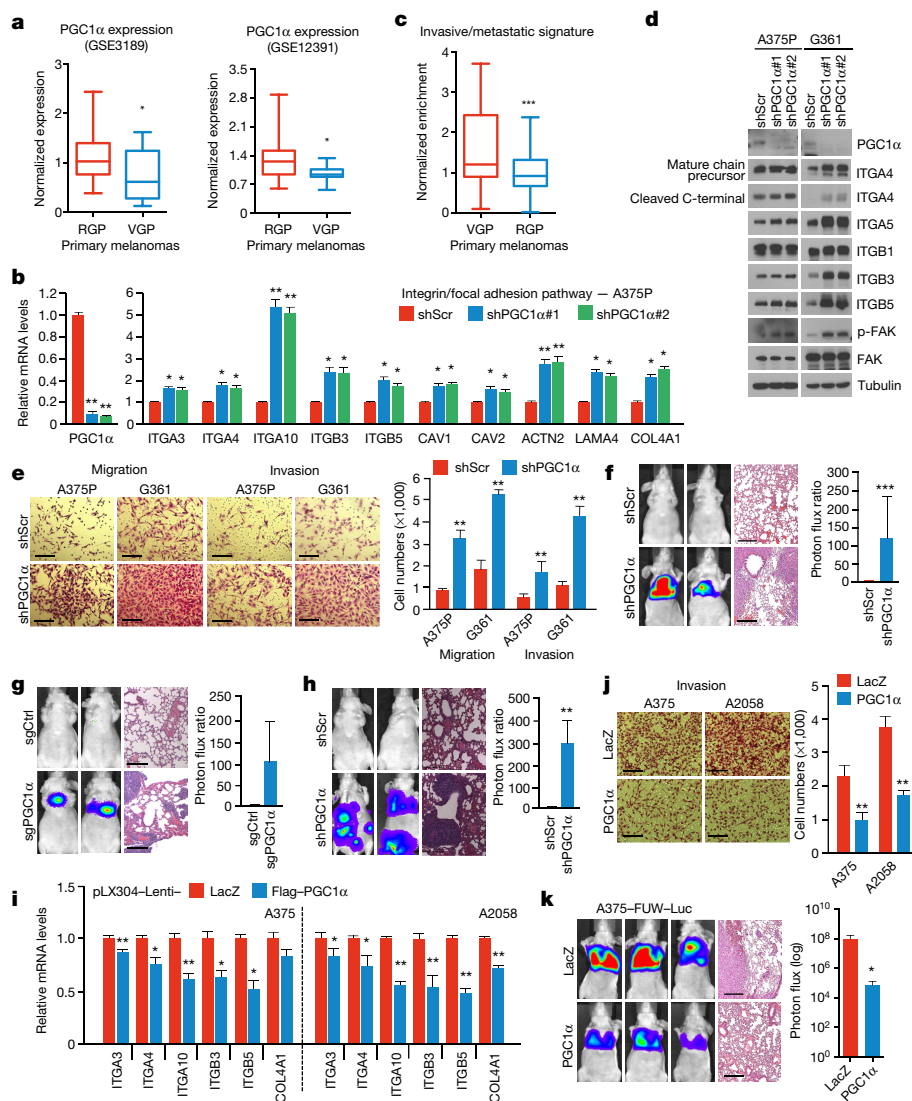


Figure 1 | PGC1 α suppresses melanoma cell migration, invasion and metastasis. **a**, PGC1 α expression is decreased in vertical growing (VGP) compared to radial growing (RGP) human primary melanomas (GSE3189: RGP $n = 36$, VGP $n = 9$; GSE12391: RGP $n = 8$, VGP $n = 15$). **b**, PGC1 α knockdown increases integrin transcripts in PGC1 α -positive melanoma cells (shScr, short hairpin scramble). **c**, Expression of the PGC1 α -regulated invasive/metastatic signature genes in human primary melanomas (VGP $n = 24$, RGP $n = 44$). **d**, PGC1 α knockdown increases integrin signalling in PGC1 α -positive melanoma cells. **e–g**, PGC1 α knockdown by shRNA (**f**, $n = 7$ mice per group) or CRISPR/Cas9 (**g**, $n = 3$ mice per group) increases melanoma cells migration/invasion (**e**) and metastasis of A375P cells (sgCtrl, small guide control) (**f**, **g**). **h**, PGC1 α knockdown elevates metastasis of subcutaneous MeWo melanoma ($n = 3$ mice per group). **i–k**, Restoration of PGC1 α suppresses integrin expression (**i**), invasion (**j**) and metastasis (**k**, $n = 4$ mice per group) of PGC1 α -negative melanoma cells. Images in **e–h**, **j** and **k** represent one picture captured, with the scale bar representing 200 μ m. Values in **a** and **c** represent median \pm relative deviation within indicated data set; values in **b**, **e**, **i** and **j** show mean \pm s.d. of independent biological triplicates; values in **f**, **g**, **h** and **k** represent mean \pm s.d. of indicated number of mice; * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.005$ by Student's t -test in all panels except **d**.

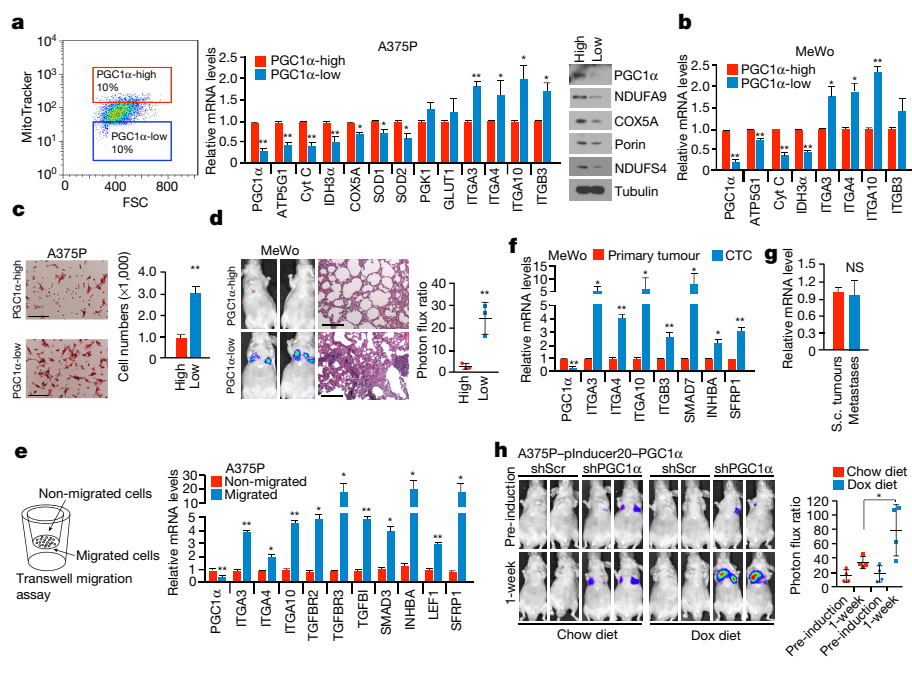


Figure 2 | The heterogeneity of PGC1 α expression in melanoma defines the metastatic capacity of individual cells.

a, Mitochondrial mass in PGC1 α -positive A375P cells correlates with PGC1 α expression. **b**, The mito/PGC1 α -low population of the PGC1 α -positive MeWo cells displays higher integrin expression. **c**, **d**, The mito/PGC1 α -low population is more migratory (**c**) and metastatic (**d**, $n = 3$ mice per group). Images represent three pictures captured with the scale bar representing 100 μ m (**c**) or 200 μ m (**d**). **e**, Within the same cell line, migratory cells express lower PGC1 α but higher pro-metastatic genes than non-migrated cells. **f**, Circulating tumour cells (CTCs) express less PGC1 α but higher pro-metastatic genes than primary tumour cells. **g**, Relative expression of PGC1 α in subcutaneous (s.c.) MeWo melanomas and lung metastases. **h**, Doxycycline (Dox)-based PGC1 α restoration in established lung metastases promotes tumour growth (shScr: $n = 3$ mice per group, shPGC1 α $n = 4$ mice per group). Quantification of tumours with shPGC1 α before and after doxycycline induction is shown. Values in **a**, **b**, **c**, **e**, **f** and **g** represent mean \pm s.d. of independent biological triplicates; values in **d** and **h** represent mean \pm s.d. of indicated number of mice; NS, not significant; * $P < 0.05$ and ** $P < 0.01$ by Student's t -test in all panels.

between PGC1 α and metabolic heterogeneity with metastatic spread, we isolated circulating tumour cells (CTCs) from the blood of mice bearing subcutaneous PGC1 α -positive MeWo tumours (Extended Data Fig. 4c). Notably, these CTCs exhibited lower levels of PGC1 α , but elevated integrins compared to the primary tumours (Fig. 2f). However, in the corresponding lung metastases, which had formed from CTCs, PGC1 α transcripts increased to similar levels as in the primary tumours (Fig. 2g). Notably, restoration of PGC1 α in lung metastases derived from PGC1 α -knockdown cells enhanced tumour progression (Fig. 2h and Extended Data Fig. 4d), further demonstrating that increases in PGC1 α in established metastases confer growth advantages similar to primary melanomas^{4,20}. In aggregate, these results indicate that melanoma cells display heterogeneous levels of PGC1 α and mitochondria. The mito/PGC1 α -low population expresses a pro-metastatic gene program, while the mito/PGC1 α -high population drives a proliferation phenotype.

To assess the mechanisms by which PGC1 α suppresses this pro-metastatic program, we surveyed genes that were upregulated upon PGC1 α suppression for potential negative transcriptional regulators. We identified two Inhibitor of DNA binding (ID) proteins—ID2 and ID3—among the top differentially expressed genes. Levels of ID2 and ID3, but not ID1 or ID4, were reduced in melanoma cells

upon PGC1 α knockdown and increased by PGC1 α (Extended Data Fig. 5a–c). Chromatin immunoprecipitation (ChIP) revealed that PGC1 α was bound at the ID2 promoter, suggesting direct transcriptional regulation (Extended Data Fig. 5d). Next, we depleted ID2 and ID3 in PGC1 α -positive melanoma cells and found that suppression of ID2, but not ID3, increased integrin expression and downstream signalling (Fig. 3a and Extended Data Fig. 5e–i). Similar to PGC1 α knockdown, ID2 depletion also strongly promoted migration, invasion and lung metastasis (Fig. 3b, c and Extended Data Fig. 5j). To test whether ID2 mediates the repressive effect of PGC1 α , we ectopically expressed ID2 in PGC1 α -depleted cells (Extended Data Fig. 6a). ID2 expression suppressed the induction of the pro-metastatic programs, invasion and metastasis enforced by PGC1 α depletion (Fig. 3d–f and Extended Data Fig. 6b, c). Similar results were observed when ID2 was ectopically expressed in PGC1 α -negative melanoma cells (Extended Data Fig. 6d–f). However, depletion of ID2, in contrast to PGC1 α , did not alter glucose metabolism (Extended Data Fig. 6g). Together, these data indicate that the ID2 inhibitor is a downstream target of PGC1 α that suppresses pro-metastatic transcriptional programs without affecting PGC1 α metabolic function.

ID2 functions as a transcriptional inhibitor through direct heterodimerization with basic helix–loop–helix (bHLH) factors, blocking

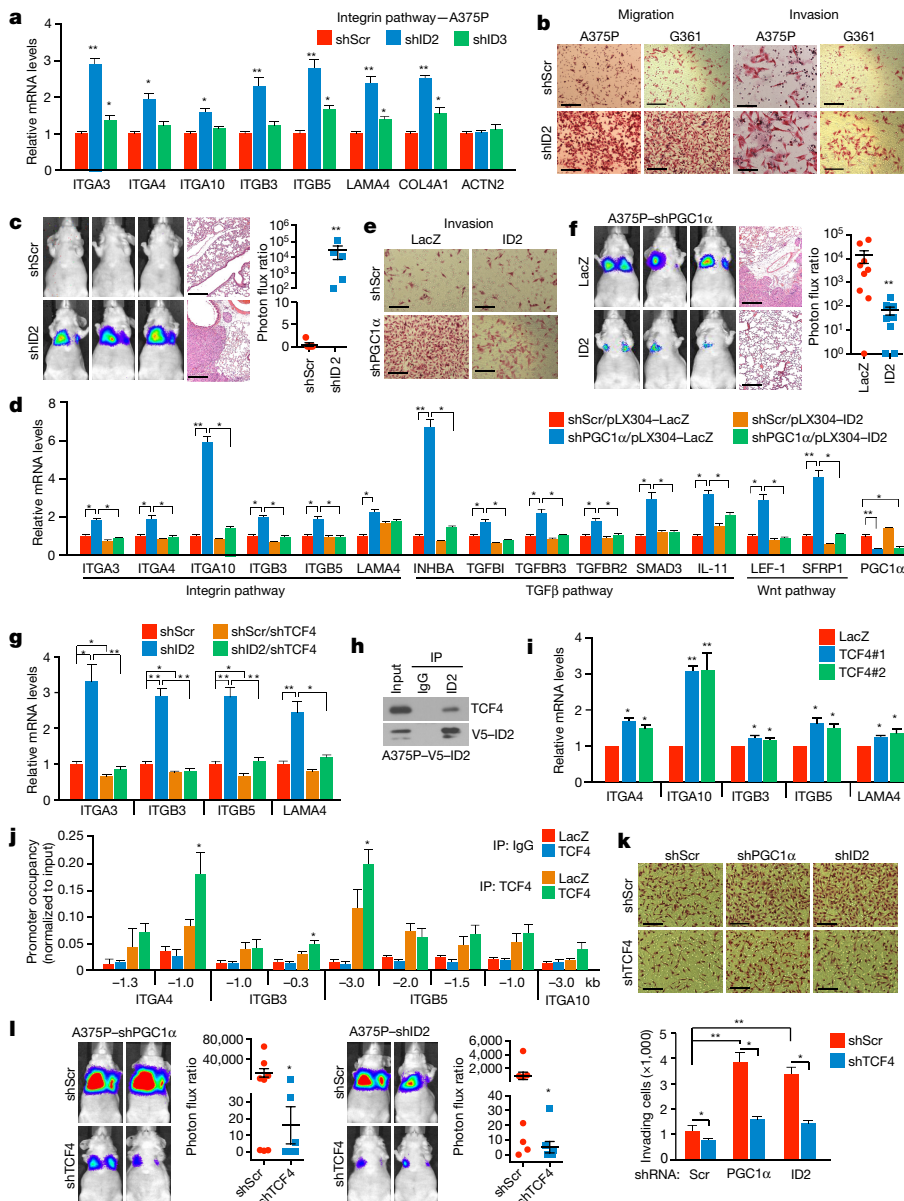


Figure 3 | PGC1 α transcriptionally activates ID2 to suppress TCF4 activity and the pro-metastatic program. **a–c**, ID2 knockdown in A375P cells increases integrin expression (**a**), migration/invasion (**b**) and metastasis (**c**, $n = 5$ mice per group). **d–f**, Ectopic expression of ID2 in A375P cells attenuates PGC1 α -depletion-mediated activation of integrin, TGF β and Wnt pathways (**d**) and induction of invasion (**e**) and metastasis (**f**, $n = 9$ mice per group). **g**, TCF4 is required for the induction of integrins by ID2 inhibition. **h**, ID2 interacts with TCF4 in A375P cells. **i, j**, Ectopic expression of TCF4 increases integrin mRNAs (**i**) through directly binding to promoters of integrin genes (**j**) in A375P cells. **k, l**, TCF4 depletion represses invasion (**k**) and metastasis (**l**, $n = 8$ mice per group) induced by PGC1 α or ID2 knockdown in A375P cells. Images in **b, c, e, f** and **k** represent one picture captured with the scale bar representing 200 μ m; specifically, scale bars in **b**, A375P invasion, represent 100 μ m. Values in **a, d, g, i, j** and **k** represent mean \pm s.d. of independent biological triplicates representative of three experiments; values in **c, f** and **l** represent mean \pm s.e.m. of indicated number of mice; in **a, c, d, f, g, i, j, k** and **l**, $*P < 0.05$, $**P < 0.01$ and $***P < 0.005$ by Student's t -test.

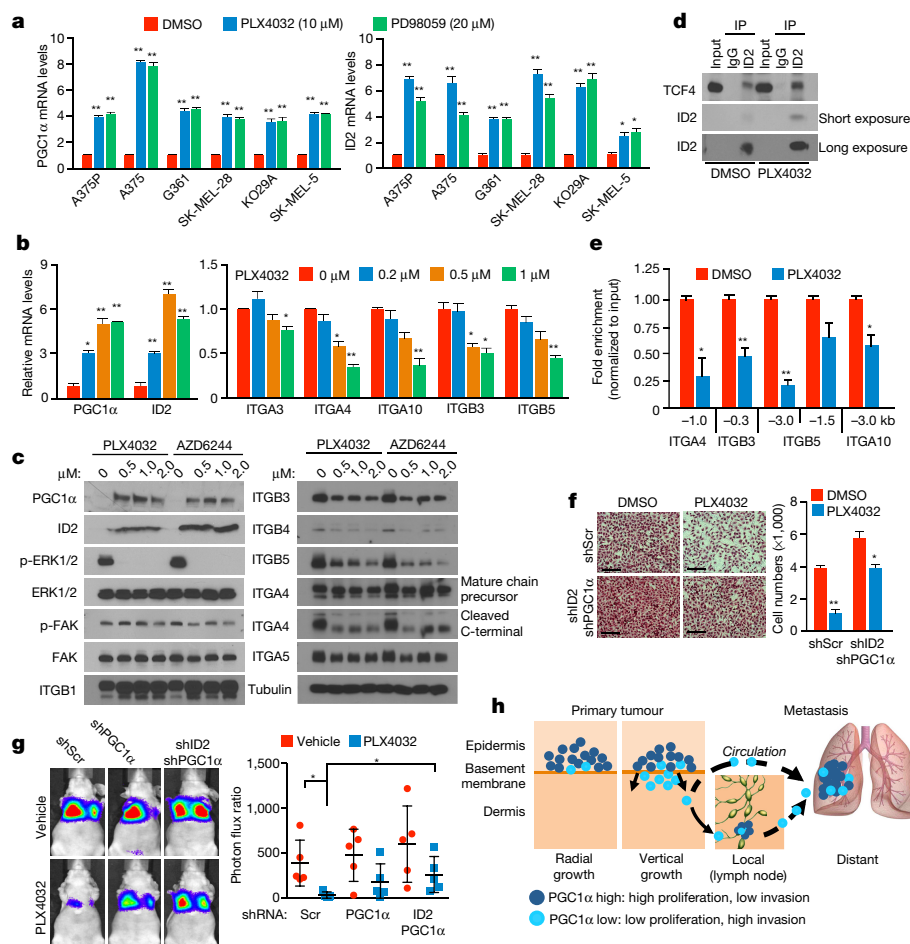


Figure 4 | BRAF^{V600E} inhibitor, PLX4032, inhibits melanoma metastasis by suppressing integrin signalling through PGC1 α and ID2.

a–c, Inhibitors of BRAF^{V600E} (PLX4032) and MEK1/2 (PD98059 or AZD6244) increase PGC1 α and ID2 expression (**a**, **b**) and repress integrin expression and signalling (**b**, **c**). Cells were treated with indicated concentration of inhibitors for 6 h (**a**, **b**) or 24 h (**c**). **d**, **e**, PLX4032 increases the interaction between ID2 and TCF4 (**d**) and decreases the occupancy of TCF4 at the promoters of integrin genes (**e**). **f–g**, PGC1 α and ID2 are partially required for PLX4032-mediated inhibition of invasion and metastasis. For *in vitro* assays (**f**), A375 cells were incubated with 1 μ M

PLX4032 for 10 h. Images represent one picture captured per membrane with the scale bar representing 200 μ m. For *in vivo* assays (**g**, $n = 5$ mice per group), PLX4032 (1 mg kg⁻¹, daily intraperitoneal injection) was given for one week following cell implantation, and metastasis was analysed three weeks post-treatment. One representative mouse image for each group is shown. **h**, Melanoma cells are heterogeneous, containing PGC1 α –high and –low subpopulations. Values in **a**, **b**, **e** and **f** represent mean \pm s.d. of independent biological triplicates; values in **g** represent mean \pm s.d. of indicated number of mice; * $P < 0.05$ and ** $P < 0.01$ by Student's *t*-test in **a**, **b**, **e**, **f** and **g**.

binding to promoters^{21–23}. To find bHLH factor(s) that could regulate integrin expression and metastasis driven by PGC1 α suppression, we surveyed two different protein–protein interaction databases. BioGRID²⁴ displayed 34 unique ID2 interactors including the bHLH transcription factors TCF3, TCF4, MyoD and TCF12 (Extended Data Fig. 7a). STRING²⁵ revealed three bHLH transcription factors (MYC, TCF3, TCF4) in the top 10 predicted partners (Extended Data Fig. 7b). Among these factors, only suppression of TCF4 was able to consistently reduce integrin expression in both A375P–shPGC1 α and PGC1 α –negative cells (Extended Data Fig. 7c–e). Knockdown of TCF4 prevented the induction of integrins and FAK phosphorylation upon PGC1 α or ID2 suppression (Fig. 3g and Extended Data Fig. 8a, b). Co-immunoprecipitation showed that TCF4 binds to ID2 in A375P cells (Fig. 3h). Consistent with this interaction, while the recruitment of TCF4 to promoters of integrins was increased upon PGC1 α or ID2 knockdown, ectopic expression of ID2 blunted TCF4 recruitment (Extended Data Fig. 8c). Ectopic expression of TCF4 was sufficient to induce integrin expression and signalling (Fig. 3i and Extended Data Fig. 8d), concordant with TCF4 recruitment to the integrin promoters (Fig. 3j). TCF4 knockdown abrogated the enhanced migration and metastasis of cells in which PGC1 α or ID2 was suppressed (Fig. 3k, l) and in PGC1 α –negative cells (Extended Data Fig. 8e). Notably,

expression of PGC1 α and TCF4 in cell lines and tumours was mutually exclusive (Extended Data Fig. 8f, g), further supporting the opposing link between PGC1 α and TCF4. Similar to ID2, manipulation of TCF4 levels did not alter cellular metabolism (Extended Data Fig. 8h), indicating that the effects of PGC1 α on metastasis are separable from its metabolic functions. Collectively, these data show that TCF4 is required for the pro-metastatic transcriptional program which leads, upon PGC1 α suppression, to increased invasion and metastasis.

PLX4032 (vemurafenib), a BRAF^{V600E} inhibitor, has been shown to increase PGC1 α expression in melanoma cells harbouring this mutation^{5,9,26}. Based on the results described here, PLX4032 could inhibit metastasis by acting on the PGC1 α transcriptional axis. Treatment of BRAF^{V600E} melanoma cells with PLX4032 or MEK inhibitors strongly induced PGC1 α and ID2 expression (Fig. 4a), and reduced levels of most integrins tested (Fig. 4b, c and Extended Data Fig. 9a). Consistently, PLX4032 increased the recruitment of PGC1 α to the ID2 promoter (Extended Data Fig. 9b) and strongly induced the interaction between ID2 and TCF4 (Fig. 4d), decreasing TCF4 promoter occupancy at four integrin genes (Fig. 4e). Despite FAK activation, measured as phospho-Y397-FAK levels, which was slightly inhibited upon MAPK blockade (Fig. 4c), PLX4032 was able to repress melanoma invasion *in vitro* and metastasis *in vivo*, which was largely reversed by

PGC1 α or ID2 depletion (Fig. 4f, g and Extended Data Fig. 9c). Within the time frame of the *in vitro* assay (24 h), PLX4032 did not decrease cell growth (Extended Data Fig. 9d) and the dose of PLX4032 used in mice (1 mg kg⁻¹) was lower than the dose used to induce tumour regression²⁷. Together, these data indicate that PLX4032 can suppress invasion and metastasis independent of its cytostatic effects. PLX4032-induced inhibition of metastasis is mediated, at least in part, through transcriptional activation of PGC1 α .

Our results overall indicate that the metabolic transcriptional coactivator PGC1 α is an apical regulator of melanoma progression through protection against oxidative stress, which confers survival and proliferative advantages^{4,5,20}, and suppression of cell motility, cell–cell interaction, adhesion and invasion that promotes metastatic drive. Notably, PGC1 α expression is inversely correlated with invasive growth in local disease, whereas in metastatic melanomas it is associated with worse outcomes. Although PGC1 α status defines a subset of melanomas with specific characteristics, its heterogeneous expression within tumours allows different proliferative or invasive abilities (Fig. 4h). We argue that the heterogeneity of PGC1 α levels within melanomas reflects a dualistic nature of PGC1 α function—promoting growth and survival of tumours, whilst suppressing metastatic spread. This heterogeneity might be important during melanoma progression through changes in PGC1 α in response to different signals including nutrients, and switching between survival–proliferation and invasion–metastasis. From a therapeutic standpoint, independent of the cytotoxic/cytostatic effects, our results extend the clinical benefits of BRAF^{V600E}-targeted drugs to metastasis. For melanoma treatment, BRAF^{V600E}-inhibitors may have heightened therapeutic benefits if applied at an earlier stage by inducing PGC1 α and reducing metastatic propensity. Moreover, selection of cells with lower PGC1 α may promote metastasis, such as within BRAF^{V600E} inhibitor-treated RAS-mutant melanomas or BRAF^{V600E} inhibitor-resistant melanomas²⁸, as these samples reactivate ERK/MEK signalling and reduce PGC1 α expression. Finally, targeting the components downstream of PGC1 α that drive metastasis could provide new therapeutic opportunities for melanoma and other malignancies such as prostate cancer²⁹.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 31 March 2014; accepted 8 August 2016.

Published online 31 August 2016.

- Chin, L., Garraway, L. A. & Fisher, D. E. Malignant melanoma: genetics and therapeutics in the genomic era. *Genes Dev.* **20**, 2149–2182 (2006).
- Gupta, P. B. *et al.* The melanocyte differentiation program predisposes to metastasis after neoplastic transformation. *Nat. Genet.* **37**, 1047–1054 (2005).
- Puigserver, P. & Spiegelman, B. M. Peroxisome proliferator-activated receptor- γ coactivator 1 α (PGC-1 α): transcriptional coactivator and metabolic regulator. *Endocr. Rev.* **24**, 78–90 (2003).
- Vazquez, F. *et al.* PGC1 α expression defines a subset of human melanoma tumors with increased mitochondrial capacity and resistance to oxidative stress. *Cancer Cell* **23**, 287–301 (2013).
- Haq, R. *et al.* Oncogenic BRAF regulates oxidative metabolism via PGC1 α and MITF. *Cancer Cell* **23**, 302–315 (2013).
- Plantefaber, L. C. & Hynes, R. O. Changes in integrin receptors on oncogenically transformed cells. *Cell* **56**, 281–290 (1989).
- Naba, A. *et al.* The matrisome: *in silico* definition and *in vivo* characterization by proteomics of normal and tumor extracellular matrices. *Mol. Cell Proteomics* **11**, M111 014647–1–18 (2012).
- Nguyen, D. X., Bos, P. D. & Massague, J. Metastasis: from dissemination to organ-specific colonization. *Nat. Rev. Cancer* **9**, 274–284 (2009).
- Bollag, G. *et al.* Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. *Nature* **467**, 596–599 (2010).
- Cancer Genome Atlas Network. Genomic classification of cutaneous melanoma. *Cell* **161**, 1681–1696 (2015).
- Hodis, E. *et al.* A landscape of driver mutations in melanoma. *Cell* **150**, 251–263 (2012).

- Fidler, I. J. The pathogenesis of cancer metastasis: the ‘seed and soil’ hypothesis revisited. *Nat. Rev. Cancer* **3**, 453–458 (2003).
- Hoek, K. S. *et al.* *In vivo* switching of human melanoma cells between proliferative and invasive states. *Cancer Res.* **68**, 650–656 (2008).
- Pinon, P. & Wehrle-Haller, B. Integrins: versatile receptors controlling melanocyte adhesion, migration and proliferation. *Pigment Cell Melanoma Res.* **24**, 282–294 (2011).
- Busse, A. & Keilholz, U. Role of TGF- β in melanoma. *Curr. Pharm. Biotechnol.* **12**, 2165–2175 (2011).
- Damsky, W. E. *et al.* β -catenin signaling controls metastasis in Braf-activated Pten-deficient melanomas. *Cancer Cell* **20**, 741–754 (2011).
- Foley, C. J. *et al.* Matrix metalloproteinase-1a promotes tumorigenesis and metastasis. *J. Biol. Chem.* **287**, 24330–24338 (2012).
- Parsons, J. T. Focal adhesion kinase: the first ten years. *J. Cell Sci.* **116**, 1409–1416 (2003).
- Pencheva, N., Buss, C. G., Posada, J., Merghoub, T. & Tavazoie, S. F. Broad-spectrum therapeutic suppression of metastatic melanoma through nuclear hormone receptor activation. *Cell* **156**, 986–1001 (2014).
- Lim, J. H., Luo, C., Vazquez, F. & Puigserver, P. Targeting mitochondrial oxidative metabolism in melanoma causes metabolic compensation through glucose and glutamine utilization. *Cancer Res.* **74**, 3535–3545 (2014).
- Langlands, K., Yin, X., Anand, G. & Prochownik, E. V. Differential interactions of Id proteins with basic-helix-loop-helix transcription factors. *J. Biol. Chem.* **272**, 19785–19793 (1997).
- Norton, J. D. ID helix-loop-helix proteins in cell growth, differentiation and tumorigenesis. *J. Cell Sci.* **113**, 3897–3905 (2000).
- Zebdree, Z. & Hara, E. Id proteins in cell cycle control and cellular senescence. *Oncogene* **20**, 8317–8325 (2001).
- Chatr-Aryamontri, A. *et al.* The BioGRID interaction database: 2013 update. *Nucleic Acids Res.* **41**, D816–823 (2013).
- Franceschini, A. *et al.* STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* **41**, D808–815 (2013).
- Luo, C. *et al.* Loss of ARF sensitizes transgenic BRAFV600E mice to UV-induced melanoma via suppression of XPC. *Cancer Res.* **73**, 4337–4348 (2013).
- Yang, H. *et al.* RG7204 (PLX4032), a selective BRAFV600E inhibitor, displays potent antitumor activity in preclinical melanoma models. *Cancer Res.* **70**, 5518–5527 (2010).
- Sanchez-Laorden, B. *et al.* BRAF inhibitors induce metastasis in RAS mutant or inhibitor-resistant melanoma cells by reactivating MEK and ERK signaling. *Sci. Signal.* **7**, ra30 (2014).
- Torrano, V. *et al.* The metabolic co-regulator PGC1 α suppresses prostate cancer metastasis. *Nat. Cell Biol.* **18**, 645–656 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank R. Bronson for his critical analysis of the mouse histology, the Nikon Imaging Center at Harvard Medical School for help with light microscopy and members of the Puigserver laboratory for discussions. J.-H.L. was supported in part by a postdoctoral fellowship from the American Heart Association (13POST14750008) and the National Research Foundation from the South-Korean government (2015R1A2A2A01002483). These studies were funded in part by the Claudia Adams Barr Program in Cancer Research (to P.P.), Dana-Farber Cancer Institute internal funds (to P.P.) and NIH R01CA181217 (to P.P.), as well as the Friends of Dana-Farber Award (to C.L.).

Author Contributions P.P. and F.V. conceived the project. C.L. contributed to immunoblots for Figs 1d, 2a, 3h, 4d, 4c and Extended Data Figs 2g, 5e, 5i, 6a, 7c, 8d, and 8f; qPCR for Figs 2, 3j, Extended Data Figs 2g, 4, 5e, 5i, and 8b; *in vitro* migration for Fig. 2c, Extended Data Figs 2h, 2i; *in vivo* experiments for Figs 1g, 1h, 2, and Extended Data Fig. 3. J.-H.L. contributed to all the other western blots, qPCR for Figs 1, 3 and 4, Extended Data Figs 2, 5, 6, 7, 8, and 9; all the *in vitro* migration and invasion experiments except Fig. 2c, Extended Data Figs 2h, 2i; all the *in vivo* experiments except Figs 1g, 1h, 2, and Extended Data Fig. 3. Y.L. contributed to all *in vivo* experiments and edited the manuscript. A.T. contributed to immunoblotting experiments for Figs 1d, 2a, 3h, 4d and 4c. F.V. and H.R.W. designed and performed the bioinformatic analyses for Extended Data Figs 1, 8g, and Fig. 1a and 1c, respectively. S.R.G. performed immunohistochemistry experiments. P.P., F.V., J.-H.L., H.R.W. and C.L. prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.P. (pere_puigserver@dfci.harvard.edu).

Reviewer Information *Nature* thanks G. Bollag and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Reagents and antibodies. PLX4032, PD98059, AZD6244 and PF-573228 were purchased from Selleck Chemicals. The siRNAs against TCF4 (sc-61657), c-Myc (sc-29226), TCF3 (sc-36618) or TCF12 (sc-35552) were purchased from Santa Cruz Biotechnology. Antibodies against ITGA4, ITGA5, ITGB1, ITGB3, ITGB4, ITGB5, FAK, c-Myc, TCF12, ERK1/2, pERK1/2 and Porin were purchased from Cell Signaling Technology; p-FAK (Y397) antibodies were from Cell Signaling and Thermo Fisher Scientific; ID2 antibodies were from Cell Signaling, Santa Cruz Biotechnology and Thermo Fisher; tubulin, V5, HMB45, COX5 A, NDUFS4 and NDUFA9 antibodies were from Abcam; TCF4 antibodies were from Abnova and Santa Cruz; and PGC1 α antibodies were from Santa Cruz and Millipore.

GSEA analysis. The GSEA software v2.0 (<http://www.broadinstitute.org/gsea>)³⁰ was used to perform the GSEA analysis. In all the analysis, the KEGG gene sets were used. The values of the 219195_at probe (corresponding to PPARGC1A) were used as phenotype. For the analysis of the CCLE data set, the gene expression data was downloaded from the CCLE portal (www.broadinstitute.org/CCLE) and the data from 61 melanoma cell lines were used in the analysis. The GSEA default parameters were used with the exception that Pearson correlation was computed to rank the genes for the analysis of the CCLE data and permutation was changed to gene set for the analysis of the GSE36879 data set.

Expression Data set Analysis. Published data sets GSE3189³¹ and GSE12391³² with associated pathological stages for each sample as Invasive/Vertical or Superficial/Radial were analysed for relative deviation from median-normalized PGC1 α intensities (linear) within each data set (significance based on 2-sample, 2-sided *t*-test statistics). To examine the enrichment of the PGC1 α -regulated metastasis/invasion signature genes (ITGA3, ITGA4, ITGA10, ITGB3, ITGB5, CAV1, CAV2, ACTN2, LAMA4A, COL4A1, INHBA, TGFBI, TGFBR3, TGFBR2, SMAD3, SMAD7, IL8, IL11, LEF1, TCF7L2, DKK3, PPP3CA and SFRP1), we performed ssGSEA projections³³ to yield a percentile-based normalized enrichment score within each of GSE3189 and GSE12391, which were used to combine the data sets (2-sample, 2-sided *t*-test statistics). The association between primary melanoma survival and PGC1 α -regulated metastasis/invasion signature was based on ssGSEA for signature closeness within GSE57715 and calculation of log-rank survival.

For the analysis of PGC1 α and TCF4 gene expression, data obtained from the TCGA skin cutaneous melanoma data set³⁴ consisting of 471 samples with RNA-seq data was downloaded from the cBioPortal^{35,36} (www.cbioportal.org). Data were represented as Z-scores of RNA-seq V2 RSM. The dotted lines denote Z-scores of 0. Samples were classified as expressing if Z-score > 0 and a mutually exclusivity report from the cBioPortal was generated.

Generation of lentiviral vectors. The pDONR223–LacZ entry control vector was purchased from Addgene (25893) and the pLX304–LacZ control vector was generated using LR clonase II (Invitrogen). The V5-tagged pLX304–ID2 and –TCF4 vectors were provided by the Marc Vidal Laboratory at Dana-Farber Cancer Institute. Luciferase-expressing FUW–Luc was provided by A. Kung and the pMSCV–Luciferase–hygro plasmid was purchased from Addgene. Full-length PGC1 α was amplified by KOD polymerase (F: GCTTGGGACATGTGCAGCGAA and R: TTACCTGCGCAAGCTTCTCTGAGC), and then the PCR product was ligated into pDONR223 by BP reaction. PGC1 α expressing destination vectors (pLX304 for constitutive expression and pInducer20³⁷ for doxycycline-inducible expression) were generated by LR reaction with entry vector (pDONR223–PGC1 α).

Cell culture, siRNA transfection, shRNA transduction and CRISPR generation. Melanoma cells were obtained from ATCC and their authentication was confirmed by either DNA fingerprinting with small tandem repeat (STR) profiling or in-house PCR testing of melanoma marker genes and BRAF mutation status. Mycoplasma contamination was tested negative in-house with the PCR Mycoplasma Detection Kit (Lonza). Melanoma cells were cultured in high-glucose DMEM containing 10% FBS. For detachment culture conditions, cells were plated on plates coated with poly-2-hydroxy methacrylate (poly-HEMA). Lentiviruses encoding shRNAs or cDNAs were produced in HEK293T cells with packaging vectors (pMD2G and psPAX2) using Polyfect (Qiagen). pLKO.1 vector expressing a scramble sequence, as listed in the Supplementary Information, was used as control (shScr). Lentiviruses particles were collected 48 h post-transfection and used to infect melanoma cells in the presence of 8 μ g/ml polybrene. Infected cells were selected with 2 μ g/ml of puromycin or 7 μ g/ml blasticidin for 4 days before experiments. siRNA transfection was performed using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Guide-RNAs were cloned into pLX–sgRNA (Addgene #50662 for PGC1 α) or lentiCRISPR (Addgene #52961 for ID2). The respective empty vector lacking the sgRNA sequence was used as control (sgCtrl). Cells were subsequently infected with lentiviruses encoding Cas9 (pCW–Cas9, Addgene #50661) and sgRNAs followed by selection with respective antibiotics as described above, and 1 μ g/ml of doxycycline for 7 days.

Western Blot. Cells were lysed in a buffer containing 1% IGEPAL, 150 mM NaCl, 20 mM HEPES (pH7.9), 10 mM NaF, 0.1 mM EDTA, 1 mM sodium orthovanadate

and 1 \times protease inhibitor cocktail. Protein concentration was quantified using the BCA protein concentration assay kit (Pierce). Cell lysates were electrophoresed on SDS-polyacrylamide gels and transferred to Immobilon-P membranes (Millipore). Membranes were incubated with primary antibodies in 5% bovine serum albumin containing 0.05% Tween-20 overnight at 4 °C. The membrane was then incubated with HRP-conjugated secondary antibody for 1 h at room temperature, and visualized using an ECL Prime (GE Healthcare).

Quantitative real-time PCR. Total RNA was isolated with Trizol (Invitrogen) by Direct-zol RNA MiniPrep kit (Zymo Research), and 2 μ g of total RNA was used for cDNA synthesis using a high capacity cDNA reverse transcription kit (Applied Biosystems). qPCR was carried out using SYBR Green PCR Master Mix (Applied Biosystems). Experimental Ct values were normalized to 36B4 where not otherwise indicated and relative mRNA expression was calculated. Sequences for all the primers are provided in the Supplementary Information.

For PGC1 α overexpression by adenovirus, A375P–shPGC1 α and A375 cells were infected with adenoviruses expressing GFP or Flag–PGC1 α for 36 h, followed by qPCR. PGC1 α targets such as GSTM4 and COX5 A were used as positive controls. For inhibitor treatment, cells were incubated with indicated concentrations of inhibitors for 6 h and mRNAs were analysed by qPCR. For the RNA from migratory and non-migratory cells, migration of the A375P and G361 cells was initiated as described below. The non-migratory cells in suspension in the upper chambers were collected by centrifugation and resuspension in lysis buffer from the Cells-to-cDNA II kit (Invitrogen). The migratory cells were collected by directly applying the lysis buffer to the membrane, following the wash and clearing of the non-migratory cells in the upper chambers. 18S rRNA was used as internal control. For cells from paraffin-embedded tissue sections, Pinpoint Slide RNA Isolation System II (Zymo Research) was used to extract RNAs.

Cell Sorting. Cells with different mitochondrial contents were sorted based on the labelling of MitoTracker Green (Invitrogen). Briefly, MitoTracker Green was spiked in the medium of 100% confluent melanoma cells at the final concentration of 75 nM, and incubated with the cells for 20 min, followed by FACS sorting at DFCI Flow Cytometry Core. The top 10% cell population with the highest mitochondrial contents (mito/PGC1 α –high) and the bottom 10% cell population with the lowest mitochondrial contents (mito/PGC1 α –low) were collected for qPCR, western blot, migration assay (1 \times 10⁵ per well overnight) and metastasis assay.

For the circulating tumour cells, whole blood of the tumour-bearing mice was collected by cardiac perfusion with PBS containing 0.5 mM EDTA. After red blood cell lysis, the pelleted cells were stained with anti-mouse CD31 and CD45, along with anti-human HLA (eBioscience, as depicted in Extended Data Fig. 4c). The CD31⁺CD45⁺hHLA⁺ cells were directly sorted into RNeasy Protect Cell Reagent (Qiagen), and then converted into cDNA using the Cells-to-cDNA II kit. The primary tumours were subjected to enzymatic digestion for single cell suspension and FACS sorting to make them equivalent controls. qPCR was performed with SYBR Green, following the unbiased, target-specific preamplification of cDNA using SsoAdvanced PreAmp Supermix (BioRad). Experimental Ct values were normalized to 18S rRNA, and relative mRNA expression was calculated.

Glucose consumption and lactate production assays. Lactate and glucose assay kits (BioVision Research Products) were used to measure extracellular lactate and glucose, following manufacturer's instructions. Briefly, equal number of cells were seeded in 6-well plates and cultured in Phenol-Red-free DMEM for 24 h or 36 h. Cultured medium was then mixed with the reaction solution. Lactate and glucose levels were measured at 450 nm and 570 nm, respectively, using a FLUO star Omega plate reader. Values were normalized to cell number.

In vivo metastasis assays. Melanoma cell lines were lifted by 0.5 mM EDTA in PBS and washed with 1 \times PBS. For the intravenous injection, a total of 3 \times 10⁵ (A375) or 1 \times 10⁶ (G361 and MeWo) or 2 \times 10⁶ (A375P and FACS-sorted MeWo) cells in 0.2 ml of DMEM were injected into the tail veins of 6-week-old male nude mice. No randomization or blinding techniques were applied in this study. To assess the degree of tumour formation in the lung, bioluminescence imaging of living mice was performed on a Xenogen IVIS-50TM imaging system equipped with an isoflurane (1–3%) anaesthesia system and a temperature-controlled platform¹⁹, three weeks (G361 and MeWo) or four weeks (A375 and A375P) post-injection. For the doxycycline-induction experiment, upon detection of lung metastasis following tail-vein implantation, PGC1 α expression was induced by feeding mice with a chow or doxycycline-containing diet (200 mg/kg, Harlan Laboratories) for one week. For the orthotopic metastasis model, 1 \times 10⁶ cells were injected subcutaneously into one side of 6-week-old male NOD/SCID mice, with two injections per animal, followed by surgical tumour removal when the subcutaneous tumours reached 2 mm in diameter. Metastasis was monitored by *in vivo* imaging at 8–10 weeks post-surgery. After the measurement of bioluminescence, animals were killed and the lungs were removed. Collected lung tissues were fixed in 10% buffered formalin solution (Sigma-Aldrich) overnight. Fixed tissues were stained with haematoxylin and eosin (H&E) or antibodies against p-FAK-Y397 (Invitrogen) or HMB45

(Abcam), and one image per sample was shown as representative of one or three pictures captured, as indicated specifically in corresponding figure legend. Scale bar represents 200 μm unless otherwise indicated. All procedures were conducted in accordance with the guidelines of the Beth Israel Deaconess Medical Center Institutional Animal Care and Use Committee, and none of the tumours exceeded the size limit dictated by the IACUC guidelines.

In vitro migration and invasion assays. For cell migration assays, transwell chambers were purchased from Corning Life Science. Generally, A375P (5×10^3) or G361 (4×10^3) cells in 0.1 ml of FBS-free medium were seeded into the upper chamber and incubated for 6 h if not otherwise indicated. For invasion assays, A375P (5×10^3), A375 (3×10^3), G361 (4×10^3), A2058 (3×10^3), RPMI7951 (5×10^3) or WM115 (4×10^3) cells in 0.1 ml of FBS-free medium were seeded into the upper chamber of an 8 μM matrigel coated chamber (BD Bioscience) and incubated for 16 h if not otherwise indicated. Specifically, for the migration and invasion assays on sorted cells (Fig. 2c) or A375P cells with ID2 knockdown (Fig. 3b), 1×10^5 cells were seeded and incubated for 24 h. Cells that had migrated and invaded through the matrigel were then fixed and stained with H&E if not otherwise indicated. The membrane attached with migrated and invaded cells was placed on a glass slide and total cell numbers from three or four random fields under 20–40 \times magnifications were quantified with an Olympus IX51 or a Nikon 80i Upright microscope, by counting cells on 20–50% of one field area and extrapolated to 100% of the field¹⁷.

Specifically for the experiments with FAK inhibitor, shScr or shPGC1 α stably expressing cells (A375P 1×10^5 per well, G361 2.5×10^4 per well) were cultured in transwell chambers with either DMSO or indicated concentration of PF-573228, followed by staining with Crystal Violet and quantification after migration for 24 h. For the experiments with PLX4032, cells were incubated with DMSO or 1 μM PLX4032 for 10 h in matrigel-coated transwell chambers, followed by quantification.

Co-immunoprecipitation and chromatin immunoprecipitation assays. Nuclear lysates were incubated with specific antibodies overnight at 4 °C, followed by precipitation with protein G Dynabeads (Invitrogen) at 4 °C for 2 h. For Fig. 3h, nuclear lysates from V5–ID2 stably-expressing A375P cells were subjected to co-IP with 1 μg ID2 antibody (C-20, Santa Cruz Biotechnology), followed by western blot with TCF4 (M03, Abnova) and ID2 (4E12G5, Thermo Scientific); for Fig. 4d, 10 mg of nuclear lysates from A375P cells treated with DMSO or 1 μM PLX4032 for 16 h were subjected to co-IP with 4 μg ID2 antibody (C-20). ChIP was performed with the MilliPore ChIP Kit with slight modification. Following sonication, nuclear lysates were precleared with protein A/G-Dynabeads (Invitrogen) for 1 h. Equal amounts of precleared lysates were incubated with IgG or gene-specific antibodies (PGC1 α 4C1.3 from Millipore, or PGC1 α H-300, and TCF4 K-15 from Santa Cruz

Biotechnology) overnight, followed by precipitation with protein A/G-Dynabeads for 2 h. qPCR with SYBR Green was performed to quantify the promoter occupancy. For Fig. 4e, A375P cells stably expressing V5–TCF4 were cultured with PLX4032 at 5 μM for 16 h and followed by ChIP and qPCR.

Cell growth and survival assays. A ToxiLight Non-destructive Cytotoxicity BioAssay Kit (Lonza) was used to quantify the cytotoxic effects of the indicated compounds according to the manufacturer's instruction. The measurement of dead cells in the DMSO group was set as 1, and was used to normalize other treatment groups (Extended Data Fig. 2j). For the cell growth assay with PLX4032 (Extended Data Fig. 9d), cells were cultured with DMSO or PLX4032 for the indicated time under either attachment or detachment conditions, followed by cell counting with a haemocytometer. For detachment culture conditions, cells were plated on tissue culture plates coated with poly-HEMA.

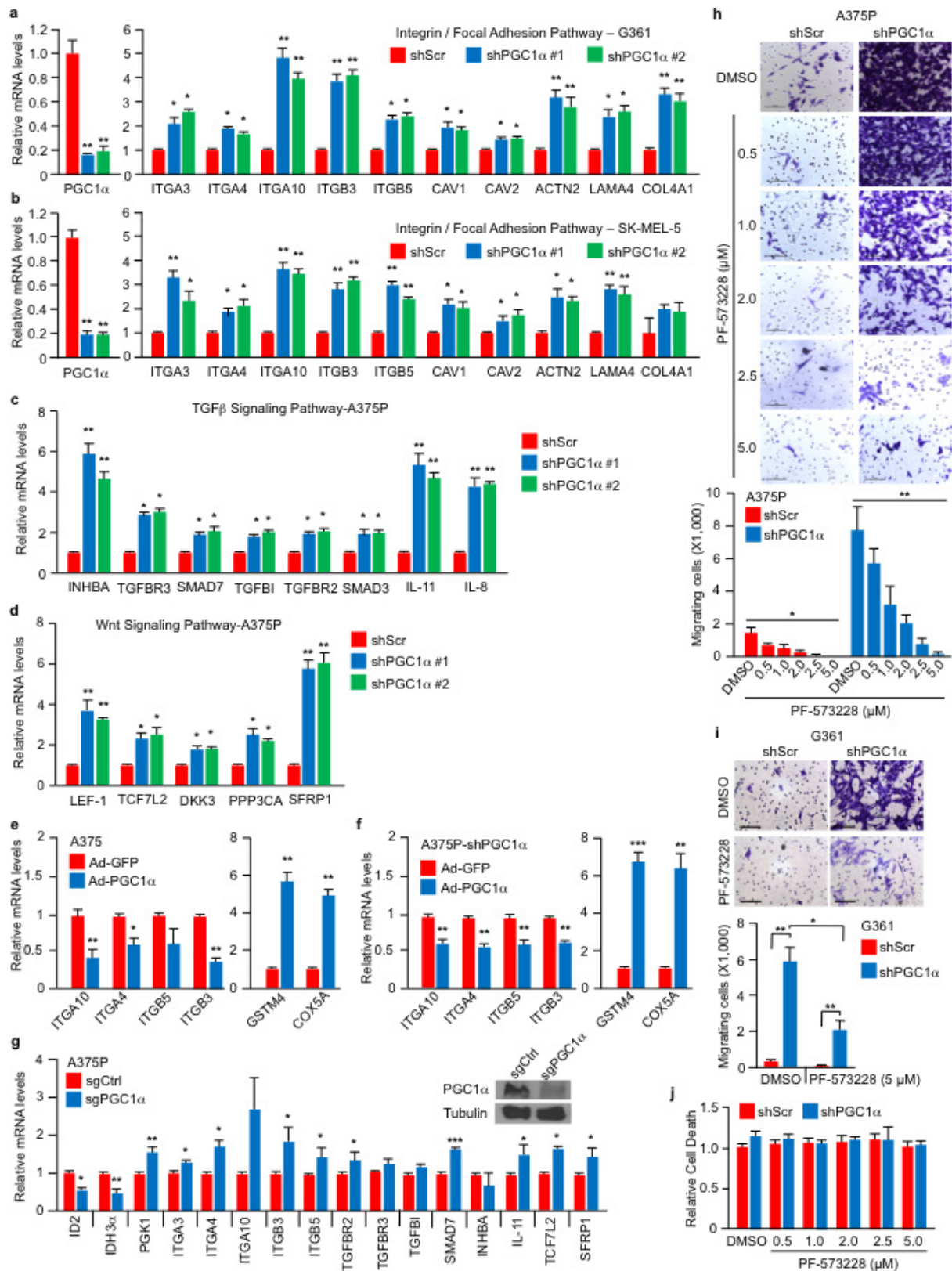
Statistics. No statistical methods were used to predetermine sample size. All statistics are described in figure legends. In general, for two experimental comparisons, a two-tailed unpaired Student's *t*-test was used unless otherwise indicated. For multiple comparisons, one-way ANOVAs were applied. When cells were used for experiments, three replicates per treatment were chosen as an initial sample size. All *n* values defined in the legends refer to biological replicates. If technical failures such as tail-vein injection failure or inadequate intraperitoneal injection occurred before collection, those samples were excluded from the final analysis. Statistical significance is represented by asterisks corresponding to **P* < 0.05, ***P* < 0.01 and ****P* < 0.005.

30. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
31. Talantov, D. *et al.* Novel genes associated with malignant melanoma but not benign melanocytic lesions. *Clin Cancer Res* **11**, 7234–7242 (2005).
32. Scatolini, M. *et al.* Altered molecular pathways in melanocytic lesions. *Int. J. Cancer* **126**, 1869–1881 (2010).
33. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).
34. Cancer Genome Atlas Network. Genomic classification of cutaneous melanoma. *Cell* **161**, 1681–1696 (2015).
35. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, p11 (2013).
36. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
37. Meerbrey, K. L. *et al.* The pINDUCER lentiviral toolkit for inducible RNA interference in vitro and in vivo. *Proc. Natl Acad. Sci. U.S.A.* **108**, 3665–3670 (2011).



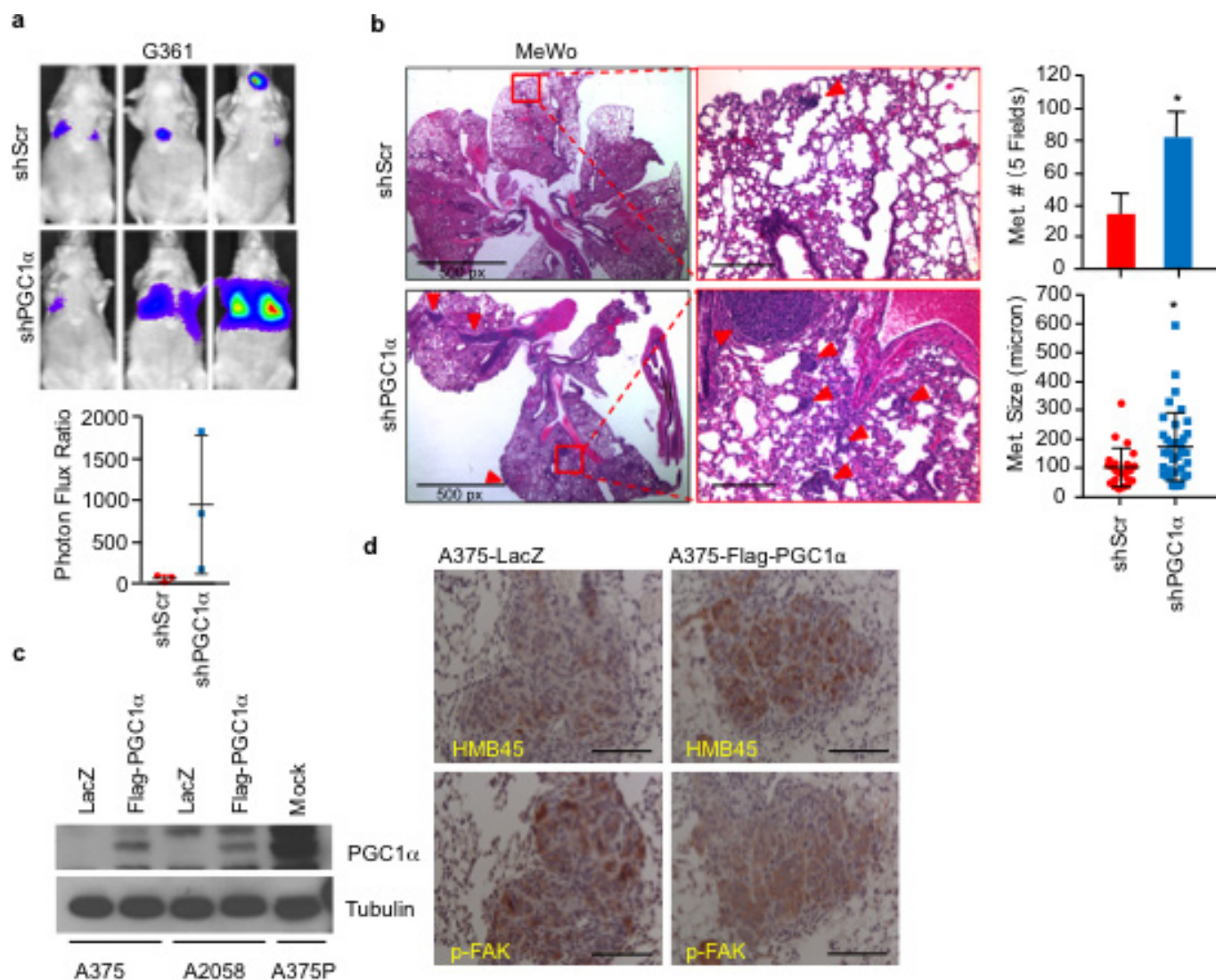
Extended Data Figure 1 | GSEA analysis of PGC1 α expression and deletion in melanoma cell lines. a, b, Representative GSEA plots (a) and list of gene sets (b) enriched in A375P cells upon PGC1 α knockdown from data set GSE36879, with the significance defined by false discovery

rate (FDR) $q < 0.25$. c, d, Plots (c) and list of (d) the top gene sets in which expression is negatively correlated with PGC1 α in 61 melanoma cell lines from CCLE, with the significance defined by FDR $q < 0.15$.



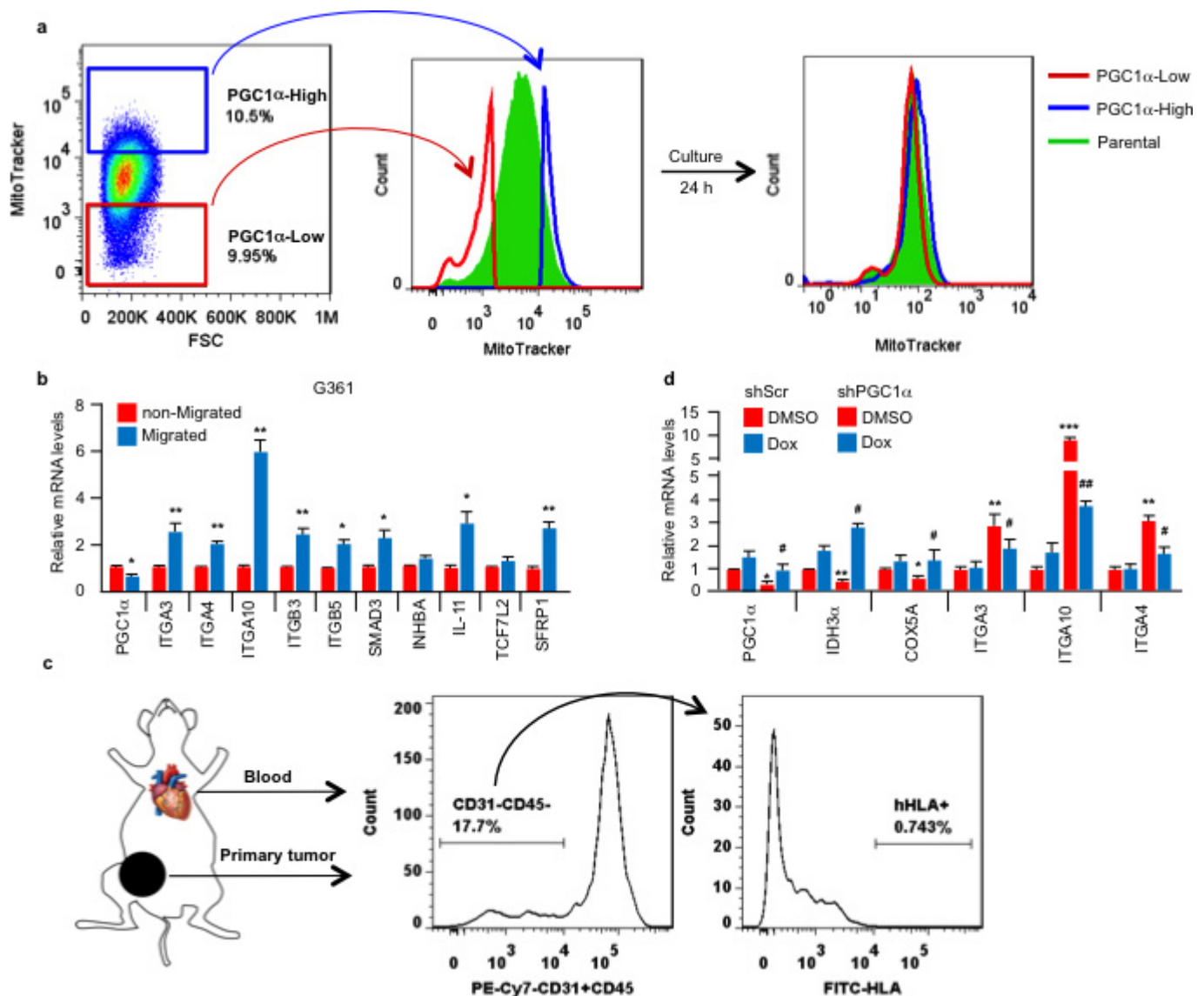
Extended Data Figure 2 | PGC1 α depletion activates integrin, TGF β and Wnt pathways. **a–d**, PGC1 α knockdown increases expression of integrin genes (**a**, **b**), as well as genes in the TGF β (**c**) and Wnt (**d**) pathways. **e**, **f**, Ectopic expression of PGC1 α by adenoviruses (Ad) inhibits integrin gene expression. **g**, CRISPR-mediated PGC1 α depletion increases gene expression linked to integrin, TGF β and Wnt pathways in A375P cells. Depletion of PGC1 α was confirmed by immunoblotting. **h**, **i**, FAK inhibition blunted the increased migration induced by PGC1 α depletion. A375P (**h**) and G361 (**i**) cells were subjected to 24 h transwell

migration assays in the presence of DMSO or various doses of FAK inhibitor PF-573228. Images represent three pictures captured with scale bar representing 100 μ m. **j**, The cytotoxic effects of the FAK inhibitor on A375P melanoma cells were comparable between the various doses used in the migration assay within the 24 h time frame. The relative level of dead cells in the culture supernatant was quantified by ToxiLight Bioassay. Values in all panels represent mean \pm s.d. of independent biological triplicates; * P < 0.05, ** P < 0.01 and *** P < 0.001 by Student's t -test in all panels.



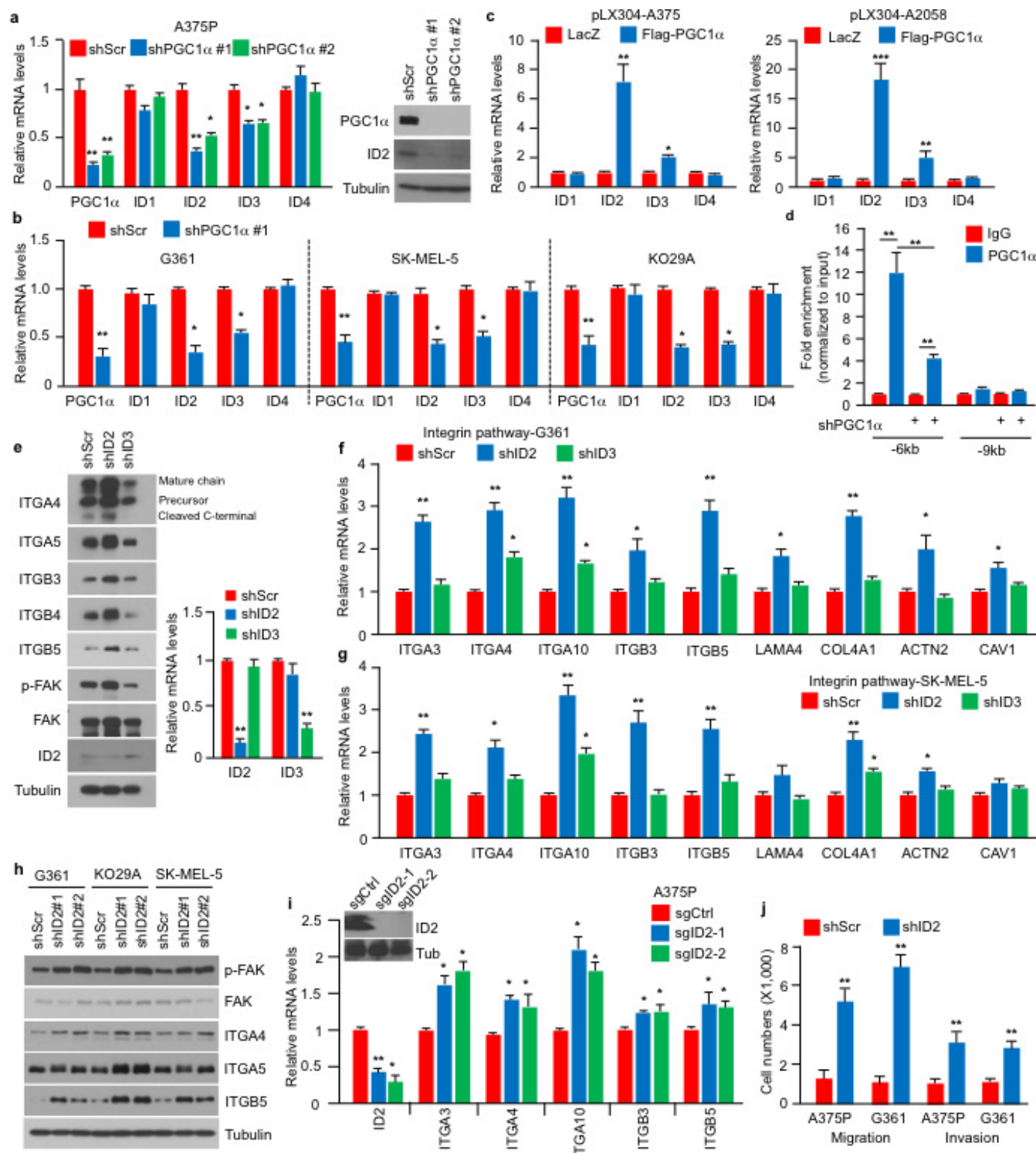
Extended Data Figure 3 | PGC1 α suppresses metastasis in melanoma cells. **a, b,** Knockdown of PGC1 α increases the metastatic capacity of PGC1 α -positive G361 (**a**, $n = 3$ mice per group) and MeWo (**b**, $n = 3$ mice per group) cells. Quantification of the number and size of lung metastatic nodules is shown. Metastatic size was quantified by measuring the longest diameter of each nodule. Values represent mean \pm s.d., $*P < 0.05$ by Student's t -test. Images in **b** represent one picture captured per H&E

slide; scale bar represents 200 μ m if not otherwise indicated. **c,** Ectopic expression level of PGC1 α in the PGC1 α -negative A375 and A2058 cell lines. **d,** Restoration of PGC1 α suppresses integrin signalling, as indicated by p-FAK (Y397), in A375-derived lung metastatic nodules. The melanoma diagnostic marker HMB45 was used to distinguish the tumour nodules from the lung tissues. Images represent three pictures captured per slide with the scale bar representing 100 μ m.



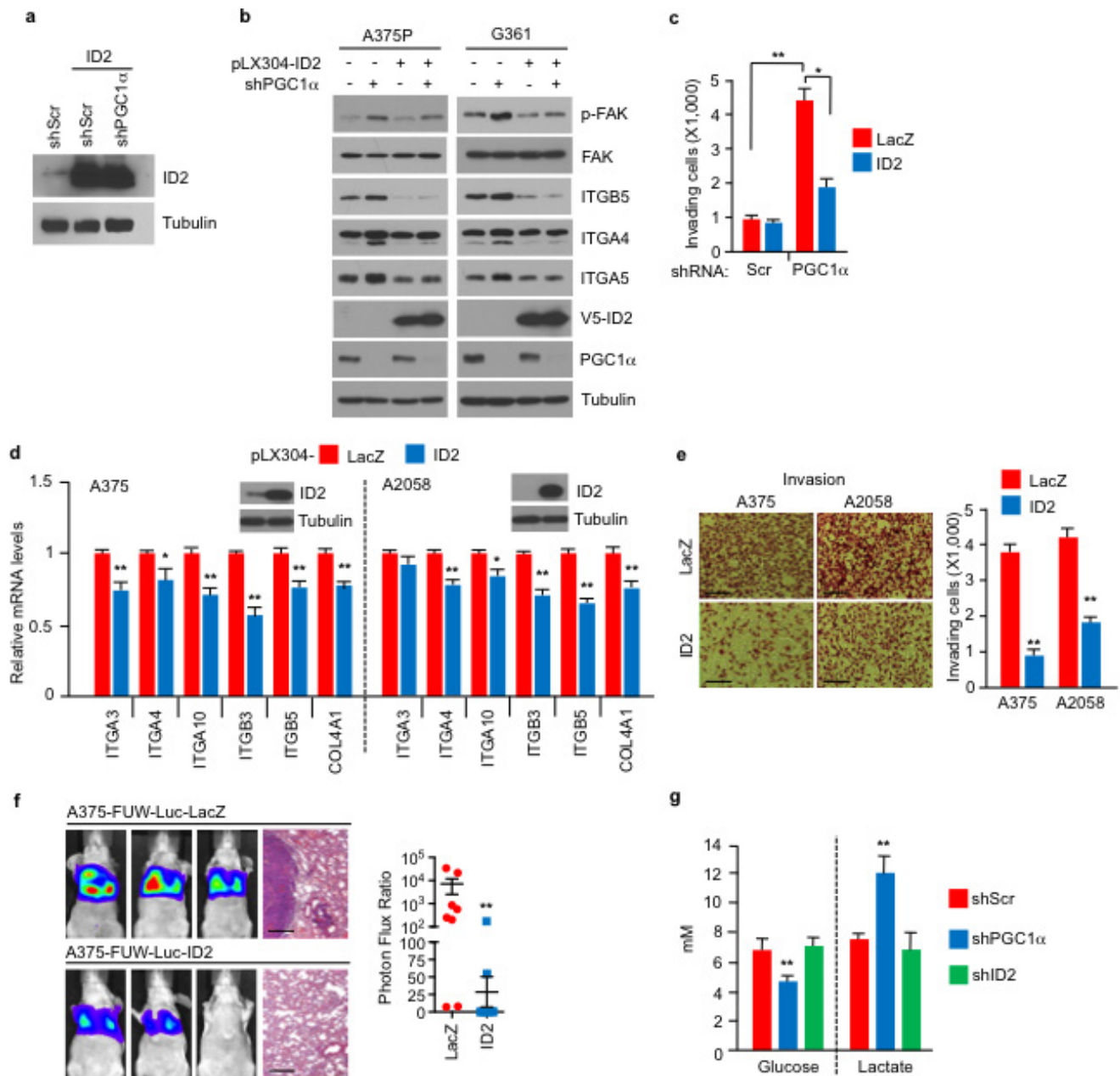
Extended Data Figure 4 | Melanoma cells contain heterogeneous levels of mitochondria and PGC1 α . **a**, The mitochondrial content in melanoma cells is dynamically regulated. After 24 h in culture, the sorted mito-high and -low A375P subpopulations re-establish normal mitochondrial content distribution. **b**, Within the PGC1 α -positive G361 line, the cells with higher migratory ability express lower PGC1 α and elevated pro-metastatic genes. Values represent mean \pm s.d. of triplicates; * P < 0.05 and ** P < 0.01 by Student's t -test. **c**, Isolation of circulating tumour cells (CTCs) from a tumour-bearing mouse. Two months post-injection, when the subcutaneous MeWo tumours became detectable, whole blood

was collected by cardiac perfusion, followed by FACS based on surface protein staining with mouse CD31 and CD45 to exclude endothelial cells and lymphocytes and human HLA^{abc} to purify human tumour cells. The primary subcutaneous tumours were enzymatically digested into single-cell suspension and subjected to the same sorting strategy. **d**, Gene expression in A375P melanoma cells after PGC1 α induction. Values represent mean \pm s.d. of independent biological triplicates; * P < 0.05, *** P < 0.01 and **** P < 0.001 versus shScr/DMSO; # P < 0.05 and ## P < 0.01 versus shPGC1 α /DMSO by Student's t -test.



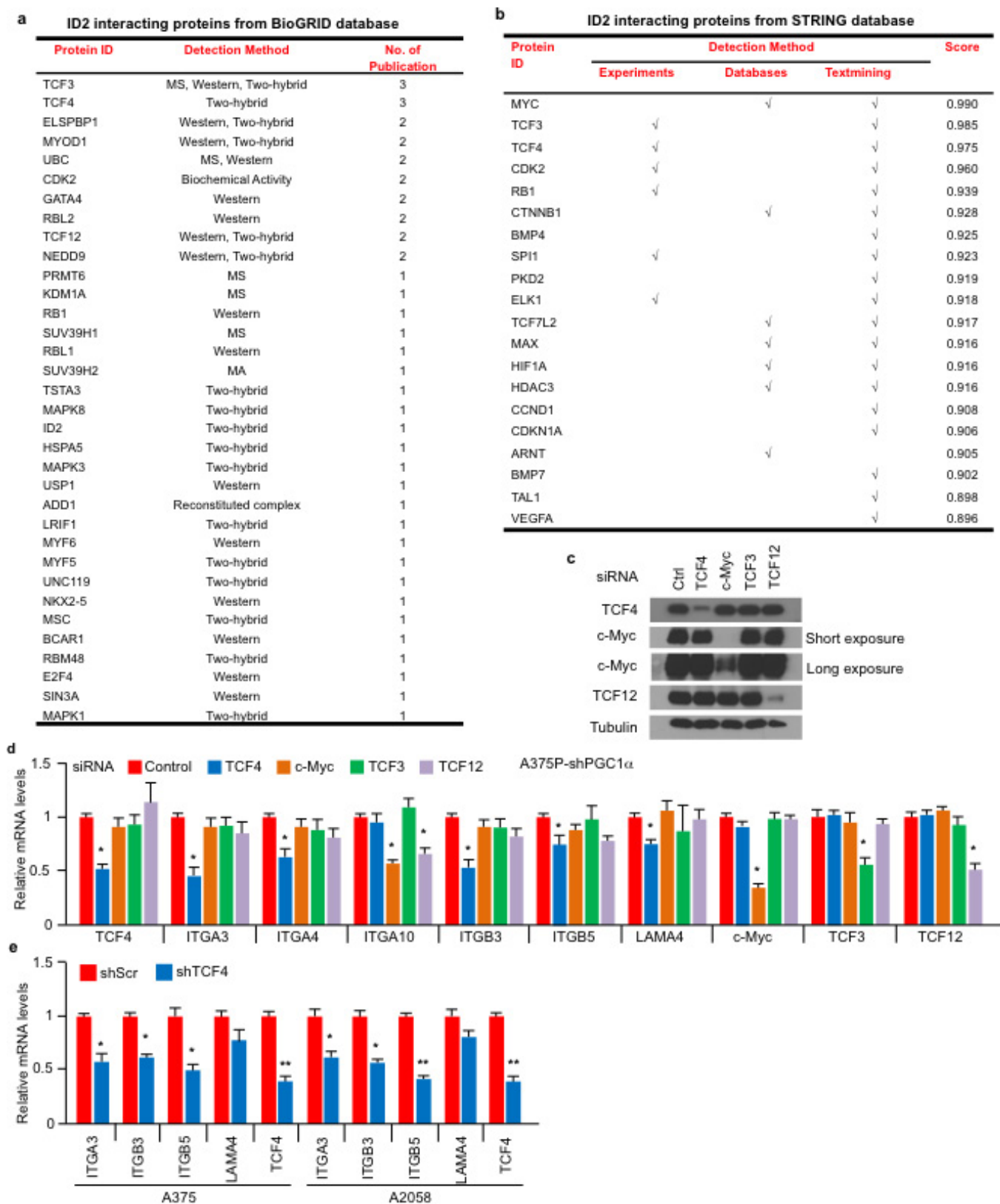
Extended Data Figure 5 | ID2, but not ID3, is downstream of PGC1α in the suppression of the pro-metastatic program. **a, b**, PGC1α knockdown inhibits ID2 expression in PGC1α-positive cells. **c**, Ectopic expression of PGC1α increases ID2 levels in PGC1α-negative cells. **d**, PGC1α occupies the ID2 promoter region in A375P cells. **e–g**, Inhibition of ID2, but not ID3, increases expression and activation of integrin

signalling in melanoma cell lines. **h, i**, Inhibition of ID2 by either shRNA or CRISPR/Cas9 increases expression of integrins. **j**, Quantification of *in vitro* migration and invasion induced by ID2 knockdown as shown in Fig. 3b. Values in all panels except **h** represent mean \pm s.d. of independent biological triplicates; * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$ by Student's *t*-test in all panels except **h**.



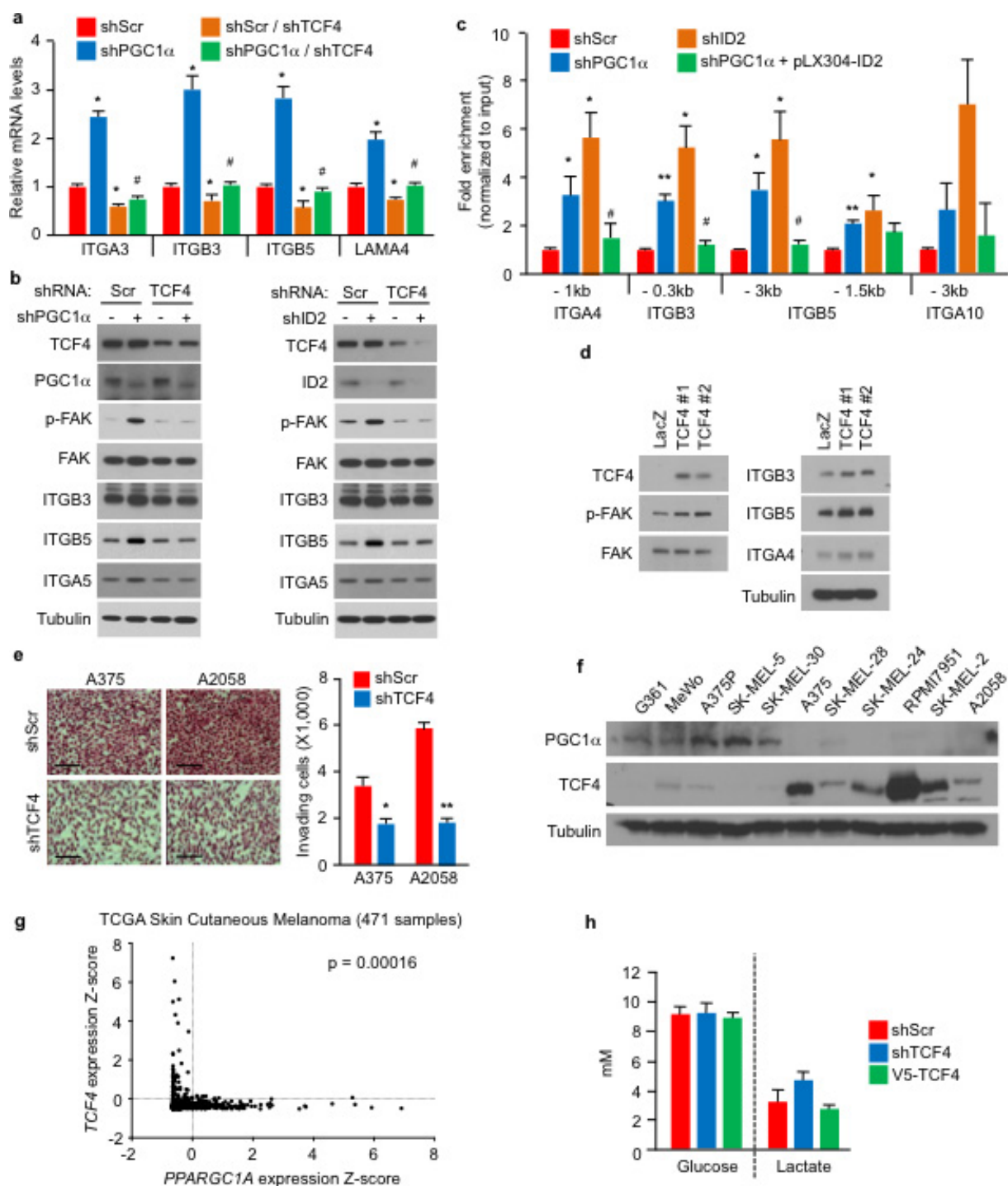
Extended Data Figure 6 | Enforced expression of ID2 suppresses metastasis. **a**, Ectopic expression of ID2 in A375P cells is higher than its endogenous level. **b**, Ectopic expression of ID2 attenuates integrin proteins and FAK (Y397) phosphorylation induced by PGC1 α depletion. **c**, Quantification of invading cells as shown in Fig. 3e. **d–f**, Ectopic expression of ID2 suppresses integrin gene expression (**d**), invasion

in vitro (**e**) and metastasis *in vivo* (**f**, $n = 8$ mice per group). Images in **e** and **f** represent one picture captured; scale bar represents 200 μ m. **g**, ID2 does not affect cellular metabolism. Values in **c**, **d**, **e** and **g** represent mean \pm s.d. of independent biological triplicates; values in **f** represent mean \pm s.e.m. of the 8 mice; * $P < 0.05$ and ** $P < 0.01$ by Student's *t*-test in **c**, **d**, **e**, **f** and **g**.



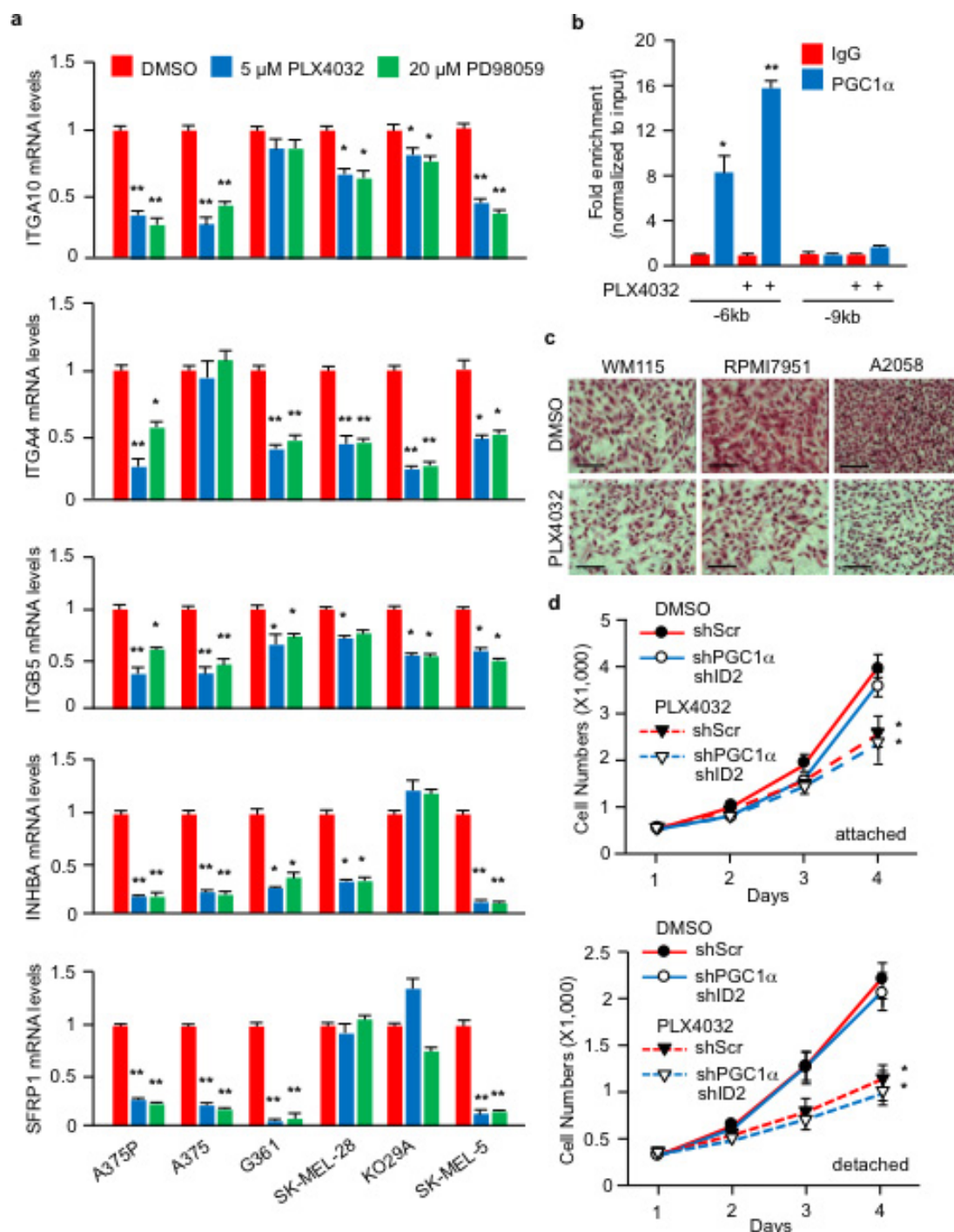
Extended Data Figure 7 | TCF4 is a putative ID2 partner in the regulation of integrin genes. **a**, **b**, List of the top ID2-interacting proteins from the BioGRID (**a**) and STRING (**b**) databases. **c**, Knockdown efficiency of individual bHLH transcriptional factors by siRNAs in A375P cells was tested by immunoblotting. **d**, Inhibition of TCF4 attenuates

PGC1 α -knockdown-mediated integrin induction in A375P cells. **e**, TCF4 knockdown suppresses gene expression linked to integrin signalling. Values in **d** and **e** represent mean \pm s.d. of independent biological triplicates; * P < 0.05 and ** P < 0.01 by Student's t -test in **d** and **e**.



Extended Data Figure 8 | TCF4 induces integrin genes. **a**, TCF4 is required for PGC1 α -depletion-mediated induction of integrin genes in A375P cells. **b**, Depletion of TCF4 blunts the activation of integrin signalling by PGC1 α or ID2 knockdown in A375P cells. **c**, Ectopic expression of ID2 blocks the binding of TCF4 to integrin promoters in A375P cells. A375P cells with indicated genetic manipulations that were stably overexpressing V5-TCF4 were subjected to ChIP and qPCR. Values in **a** and **c** represent mean \pm s.d. of independent biological triplicates; * $P < 0.05$ and ** $P < 0.01$ versus shScr; # $P < 0.05$ versus shPGC1 α by Student's *t*-test. **d**, Ectopic expression of TCF4 increases integrin proteins

and signalling in A375P cells. **e**, TCF4 knockdown suppresses cell invasion in A375 and A2058 cells. Images represent one picture captured per membrane with the scale bar representing 200 μ m. **f**, Expression of PGC1 α and TCF4 in a panel of human melanoma cell lines. **g**, TCF4 and PGC1 α expression in TCGA skin cutaneous melanoma dataset (471 samples with RNA-seq expression data). Tendency towards mutual exclusivity for samples with Z-scores > 0 (represented by dotted lines), $P = 0.00016$ by Fisher's exact test. **h**, TCF4 level does not affect cellular metabolism. Values in **e** and **h** represent mean \pm s.d. of independent biological triplicates; * $P < 0.05$ by Student's *t*-test.



Extended Data Figure 9 | BRAF^{V600E} inhibitor suppresses melanoma invasion independent of its cytostatic effect. **a**, The BRAF^{V600E} inhibitor, PLX4032, and the MEK1/2 inhibitor, PD98059, decrease integrin gene expression in melanoma cell lines. Gene expression was quantified 6 h post-treatment of inhibitors. **b**, PLX4032-induced PGC1 α occupancy at the ID2 promoter. A375P cells were incubated with 2 μ M of PLX4032 for 6 h before ChIP analysis. **c**, PLX4032 inhibits invasion of BRAF^{V600E}

containing melanoma cells. Cells were incubated with 1 μ M PLX4032 for 10 h in matrigel-coated transwell chambers, followed by quantification. Images represent one picture captured per membrane with the scale bar representing 200 μ m. **d**, PGC1 α and ID2 double knockdown does not affect sensitivity to PLX4032. Values in **a**, **b** and **d** represent mean \pm s.d. of independent biological triplicates; * P < 0.05 and ** P < 0.01 by Student's t -test in **a**, **b** and **d**.

Restricted diet delays accelerated ageing and genomic stress in DNA-repair-deficient mice

W. P. Vermeij^{1*}, M. E. T. Dollé^{2*}, E. Reiling^{1,2}, D. Jaarsma³, C. Payan-Gomez^{1,4}, C. R. Bombardieri¹, H. Wu⁵, A. J. M. Roks⁵, S. M. Botter^{1,6}, B. C. van der Eerden⁷, S. A. Youssef⁸, R. V. Kuiper^{8†}, B. Nagarajah², C. T. van Oostrom², R. M. C. Brandt¹, S. Barnhoorn¹, S. Imholz², J. L. A. Pennings², A. de Bruin^{8,9}, A. Gyenis¹, J. Pothof¹, J. Vijg¹⁰, H. van Steeg^{2,11} & J. H. J. Hoeijmakers^{1,12}

Mice deficient in the DNA excision-repair gene *Ercc1* (*Ercc1*^{Δ/Δ}) show numerous accelerated ageing features that limit their lifespan to 4–6 months^{1–4}. They also exhibit a ‘survival response’, which suppresses growth and enhances cellular maintenance. Such a response resembles the anti-ageing response induced by dietary restriction (also known as caloric restriction)^{1,5}. Here we report that a dietary restriction of 30% tripled the median and maximal remaining lifespans of these progeroid mice, strongly retarding numerous aspects of accelerated ageing. Mice undergoing dietary restriction retained 50% more neurons and maintained full motor function far beyond the lifespan of mice fed *ad libitum*. Other DNA-repair-deficient, progeroid *Xpg*^{−/−} (also known as *Ercc5*^{−/−}) mice, a model of Cockayne syndrome⁶, responded similarly. The dietary restriction response in *Ercc1*^{Δ/Δ} mice closely resembled the effects of dietary restriction in wild-type animals. Notably, liver tissue from *Ercc1*^{Δ/Δ} mice fed *ad libitum* showed preferential extinction of the expression of long genes, a phenomenon we also observed in several tissues ageing normally. This is consistent with the accumulation of stochastic, transcription-blocking lesions that affect long genes more than short ones. Dietary restriction largely prevented this declining transcriptional output and reduced the number of γH2AX DNA damage foci, indicating that dietary restriction preserves genome function by alleviating DNA damage. Our findings establish the *Ercc1*^{Δ/Δ} mouse as a powerful model organism for health-sustaining interventions, reveal potential for reducing endogenous DNA damage, facilitate a better understanding of the molecular mechanism of dietary restriction and suggest a role for counterintuitive dietary-restriction-like therapy for human progeroid genome instability syndromes and possibly neurodegeneration in general.

Dietary restriction is the best-documented intervention for extending lifespan in numerous species and retards many symptoms of ageing^{7–10}. Despite extensive research, its underlying mechanisms are still unresolved, although suppression of growth hormone/insulin-like growth factor 1 (GH/IGF1) and mechanistic target of rapamycin (mTOR) signalling are likely implicated^{8,10}. The molecular underpinnings of ageing itself are also poorly understood, although the fact that progeroid syndromes are associated with impaired genome maintenance⁴ points towards a connection with compromised genome stability^{11,12}. Links between the accumulation of DNA damage and the GH/IGF1 axis emerged when DNA-repair-deficient progeroid

mice were found to have a suppressed GH/IGF1 somatotrophic axis and upregulated anti-oxidant defences, presumably in an attempt to extend their lifespan by redirecting resources from growth to cellular maintenance and stress resistance^{1,5}. Normal mice and mammalian cells also share this ‘survival response’ after the induction of persisting DNA damage, indicating that it is a common response^{1,5,13}.

Growth-suppressed progeroid DNA-repair mutants show spontaneous dietary-restriction-like responses alongside other signs suggestive of dietary restriction, including reduced subcutaneous fat and paradoxical features of delayed ageing^{14,15}. We therefore wondered whether subjecting progeroid mice to actual dietary restriction would be beneficial or, in view of their poor growth and frail appearance, detrimental. We subjected *Ercc1*^{Δ/Δ} progeroid DNA-repair mutants, with a lifespan of only 4–6 months^{3,16,17}, to gradual food restriction. Mice initially had food restricted by 10% at week 7, with restriction reaching a maximum of 30% from 9 weeks onwards. Dietary restriction in both genders extended median and maximal remaining lifespan by approximately 200%; the median lifespan of males increased from 10 to 35 weeks (250% extension; $P < 0.0001$) and that of females increased from 13 to 39 weeks (200% extension; $P < 0.0001$) (Fig. 1a, b).

As lifespan can be influenced by factors other than food¹⁸, we repeated the study in another animal facility with different housing but similar food and a similarly restricted diet. A dietary restriction of 30% extended median remaining lifespan by 180% ($P < 0.0001$; Fig. 1c). We decided to test another repair-deficient progeroid mutant, the Cockayne syndrome-like *Xpg*^{−/−} mouse. This genotype carries defects in partially different DNA repair pathways² and has an even shorter lifespan than the *Ercc1*^{Δ/Δ} mutant (approximately 18 weeks versus 22–25 weeks)⁶. The same dietary restriction regimen induced a significant increase in remaining median lifespan of approximately 80% ($P < 0.0001$; Fig. 1d), widening the scope of dietary restriction beyond *Ercc1* mutants. Even a six-week dietary restriction interval, from 6 to 12 weeks of age, yielded a striking median lifespan extension of 6 weeks for *Ercc1*^{Δ/Δ} mice ($P = 0.0042$) and 4 weeks for *Xpg*^{−/−} mice ($P < 0.0001$; Fig. 1c, d), indicating that the effect brought about by dietary restriction persists, consistent with reducing the short-term death risk found in *Drosophila*¹⁹. When comparing *Ercc1*^{Δ/Δ} with *Xpg*^{−/−} it should be noted that *Xpg*^{−/−} animals were already biologically older when dietary restriction started. Although *Ercc1*^{Δ/Δ} and *Xpg*^{−/−} mice fed *ad libitum* never become obese, upon dietary restriction, their bodyweights uniformly stabilized to gender- and genotype-specific

¹Department of Molecular Genetics, Erasmus University Medical Center Rotterdam, PO Box 2040, 3000 CA Rotterdam, The Netherlands. ²Centre for Health Protection, National Institute for Public Health and the Environment (RIVM), PO Box 1, 3720 BA Bilthoven, The Netherlands. ³Department of Neuroscience, Erasmus University Medical Center Rotterdam, PO Box 2040, 3000 CA Rotterdam, The Netherlands. ⁴Facultad de Ciencias Naturales y Matemáticas, Universidad del Rosario, Carrera 24, 63C-69 Bogotá, Colombia. ⁵Department of Internal Medicine, Division of Vascular Medicine and Pharmacology, Erasmus University Medical Center Rotterdam, PO Box 2040, 3000 CA Rotterdam, The Netherlands. ⁶Laboratory for Orthopedic Research, Balgrist University Hospital, Forchstrasse 340, 8008, Zürich, Switzerland. ⁷Department of Internal Medicine, Erasmus University Medical Center Rotterdam, PO Box 2040, 3000 CA Rotterdam, The Netherlands. ⁸Dutch Molecular Pathology Center, Department of Pathobiology, Faculty of Veterinary Medicine, Utrecht University, PO Box 80125, 3508 TC Utrecht, The Netherlands. ⁹Department of Pediatrics, Division Molecular Genetics, University Medical Center Groningen, PO Box 30001, 9700 RB Groningen, The Netherlands. ¹⁰Department of Genetics, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, USA. ¹¹Department of Human Genetics, Leiden University Medical Center, PO Box 9600, 2300 RC Leiden, The Netherlands. ¹²CECAD Forschungszentrum, Universität zu Köln, Joseph-Stelzmann-Straße 26, 50931 Köln, Germany. [†]Present address: Department of Laboratory Medicine, Karolinska Institute, SE-171, 77 Stockholm, Sweden.

*These authors contributed equally to this work.

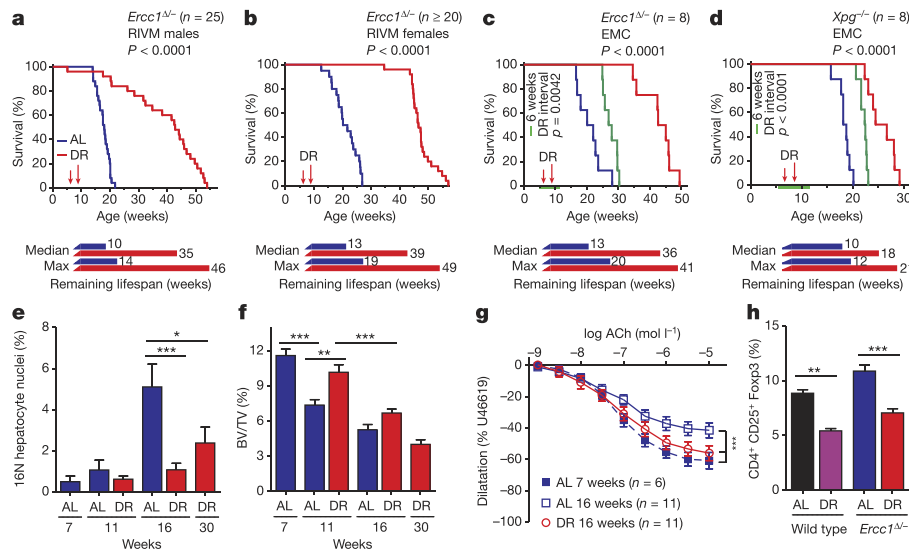


Figure 1 | Dietary restriction extends health and lifespan of *Ercc1*^{Δ/Δ} and *Xpg*^{-/-} mouse mutants. **a–d**, Survival of mice with *ad libitum* access to AIN93G diet or on 30% dietary restriction (dietary restriction, red; *ad libitum*, blue throughout) at two separate test sites. Male (**a**) and female (**b**) *Ercc1*^{Δ/Δ} mice, group housed at the National Institute for Public Health and the Environment (RIVM) (n = 20–25 animals per group, separate experiments), and *Ercc1*^{Δ/Δ} (**c**) and *Xpg*^{-/-} (**d**) mice, solitary housed at the Erasmus University Medical Center Rotterdam (EMC) (n = 8 animals per group, 4 of each gender), under *ad libitum* or dietary restriction regimens. Dietary restriction was initiated at 7 weeks of age with 10% restriction, and increased weekly by 10%, until 30% was reached from 9 weeks onward. Remaining median and maximum lifespan are indicated (week 8 was considered the start of effective dietary restriction). Simultaneously, a cohort of *Ercc1*^{Δ/Δ} (**c**) and *Xpg*^{-/-} (**d**) mice underwent temporary dietary restriction for 6 weeks (30% dietary restriction from 6 to 12 weeks)

(green; n = 8 animals per group, 4 of each gender). P values were calculated by the log-rank test. **e**, Quantification of 16N nuclei in hepatocytes of *ad libitum* or dietary restriction male *Ercc1*^{Δ/Δ} mice by FACS analyses; n = 5 animals per group. **f**, Trabecular bone volume fraction (bone volume/tissue volume of interest, BV/TV) in femurs of *Ercc1*^{Δ/Δ} male mice, measured using micro-CT. *Ad libitum*- and dietary restriction-treated animals were analysed at different ages with n ≥ 6 animals per group. **g**, Age-dependent decline of vasodilatation in *Ercc1*^{Δ/Δ} aorta segments, *ex vivo*. Dietary restriction-*Ercc1*^{Δ/Δ} aorta segments show significantly more relaxation at age of 16 weeks than *ad libitum*-*Ercc1*^{Δ/Δ} aorta. ACh, acetylcholine. **h**, Frequency of CD4⁺ CD25⁺ Foxp3⁺ T-regulatory cells among all CD4⁺ T cells from spleen of 16-week-old *Ercc1*^{Δ/Δ} mice under dietary restriction or *ad libitum* and aged-matched wild-type controls. n ≥ 3 animals per group. Error bars denote mean ± s.e. *P < 0.05, **P < 0.01, ***P < 0.001.

values, persisting until the end of their extended lifespan. Paradoxically, however, when bodyweights of mice fed *ad libitum* approached the weight of diet-restricted mice, they died (Extended Data Fig. 1a–d).

As the extension of lifespan can be unrelated to ageing¹⁸, we examined key ageing parameters. *Ercc1*^{Δ/Δ} mutants exhibit exceptionally wide multi-morbidity, consistent with the notion that this protein is implicated in multiple DNA repair processes, including both transcription-coupled and global-genome nucleotide excision repair (NER) and crosslink repair². Ageing in *Ercc1*^{Δ/Δ} mice involves proliferative and post-mitotic organs, including the nervous system, liver, kidney, bone marrow, retina, muscle and the cardiovascular, skeletal, and gonadal systems^{1–4}. Additionally, they exhibit progressively declining vision and hearing, sarcopenia, cachexia, overall frailty and a variety of other features of ageing, many of which are also noted in XpF/*Ercc1* (XFE) and other related human syndromes^{1,20,21}. Multisystem cross-sectional analyses showed that dietary restriction strongly attenuated virtually all features of premature ageing that were investigated, including anisokaryosis in the liver and kidney, formation of polyploid liver nuclei, kidney tubulonephrosis, osteoporosis, vascular dilatation, B- and T-cell immune parameters and testicular degeneration (Fig. 1e–h, Extended Data Figs 1–3 and Supplementary Table 1). Although the premature-ageing phenotype displays extreme multi-morbidity, an important cause of death of *Ercc1*^{Δ/Δ} mice is the progressive neurodegeneration that parallels NER-deficient premature ageing conditions in humans^{22,23}. Considering this translational relevance, we analysed neurological function in more detail. Longitudinal examination of behavioural abnormalities showed that the onset of tremors, imbalance, and paresis were greatly postponed or even absent in *Ercc1*^{Δ/Δ} and *Xpg*^{-/-} mice that underwent continuous or temporary dietary restriction regimes (Fig. 2a–c, Extended Data Fig. 4 and Supplementary Video 1). Dietary restriction strongly

improved motor function in *Ercc1*^{Δ/Δ} mice; at 16 weeks of age mice fed *ad libitum* displayed severe locomotor problems and frequently fell, whereas diet-restricted mice were fully capable of running (Fig. 2d and Supplementary Video 2). Even at ages far beyond the lifespan of mice fed *ad libitum*, locomotor function is well preserved in mice undergoing dietary restriction (Fig. 2e). In line with behavioural data indicating that neurological decline was virtually stopped, neurodegenerative pathology was strongly diminished by dietary restriction (including retinal photoreceptor loss, Golgi abnormalities, axonal swellings, astrocytosis and microgliosis) (Extended Data Figs 2e, 5, 6a, b). Notably, stereological counting revealed that diet-restricted animals retained approximately 50% more neurons in the neocortex than controls fed *ad libitum* (Fig. 2f), while the number of non-neuronal cells was similar (Extended Data Fig. 6c). Likewise, significantly more motor neurons were preserved in the spinal cord upon dietary restriction (Fig. 2g). These findings indicate that dietary restriction greatly improves health and lifespan in both *Ercc1* and *Xpg* repair-deficient mice and in particular attenuates neurodegeneration.

To examine whether the effects of dietary restriction in *Ercc1*^{Δ/Δ} mice resemble those in wild-type mice, we compared full-genome liver mRNA expression profiles in 11-week-old mice. Unbiased principal-component analysis of four groups: wild-type *ad libitum* (AL^{WT}), *Ercc1*^{Δ/Δ} *ad libitum* (AL^{*Ercc1*}), wild-type with 30% dietary restriction (DR^{WT}) and *Ercc1*^{Δ/Δ} with 30% dietary restriction (DR^{*Ercc1*}) revealed clear uniformity within and distinction between them, primarily based on dietary restriction and genotype (Fig. 3a). Of the 1,106 differentially expressed genes (DEGs) seen in DR^{*Ercc1*} mice, around two-thirds were also seen in DR^{WT} mice. Pertinently, expression levels of 684 out of the 688 common DEGs are also changed in the same direction in DR^{WT} and DR^{*Ercc1*} mice (Table 1 and Extended Data Table 1), indicating strong mechanistic parallels between both dietary restriction responses,

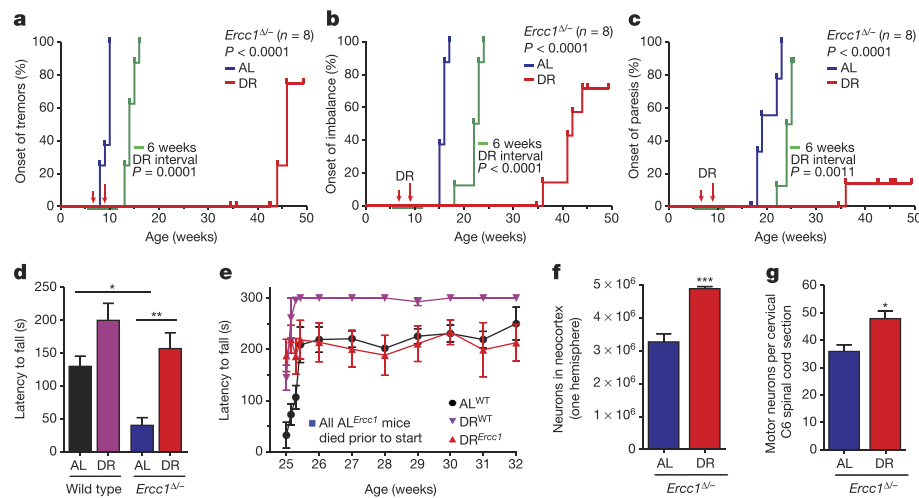


Figure 2 | Dietary restriction preserves neurological function.

a–c, Onset of neurological abnormalities (tremors (**a**), imbalance (**b**), paresis of the hind limbs (**c**)) with age in *ad libitum* and dietary restriction *Ercc1*^{Δ/Δ} mice. *n* = 8 animals per group. The onset of continuous dietary restriction is indicated by red arrows and the 6-week dietary restriction interval as green horizontal line. **d, e**, Average time spent on an accelerating rotarod of wild-type and *Ercc1*^{Δ/Δ} mice on different diets at 16 weeks of age (**d**; *n* = 8 animals per group) or weekly monitored beyond the lifespan of *AL*^{*Ercc1*} mice (**e**; *Ercc1*^{Δ/Δ} *n* = 4, wild-type *n* = 3). A daily

training period was given at 25 weeks of age. **f, g**, Quantitative stereological analysis of the total number of neurons (f, NeuN⁺; *P* = 0.0008) in the neocortex of transverse brain sections and motor neurons (**g**, ChAT⁺; *P* = 0.0176) in C6 cervical spinal cord sections of 16-week-old *ad libitum* and diet-restricted *Ercc1*^{Δ/Δ} mice. Note that the selective effect on neurons is consistent with earlier observations that neurons are the primary target of *Ercc1* deficiency²⁸. *N* ≥ 3 animals per group. Error bars indicate mean ± s.e. **P* < 0.05, ***P* < 0.01, ****P* < 0.001.

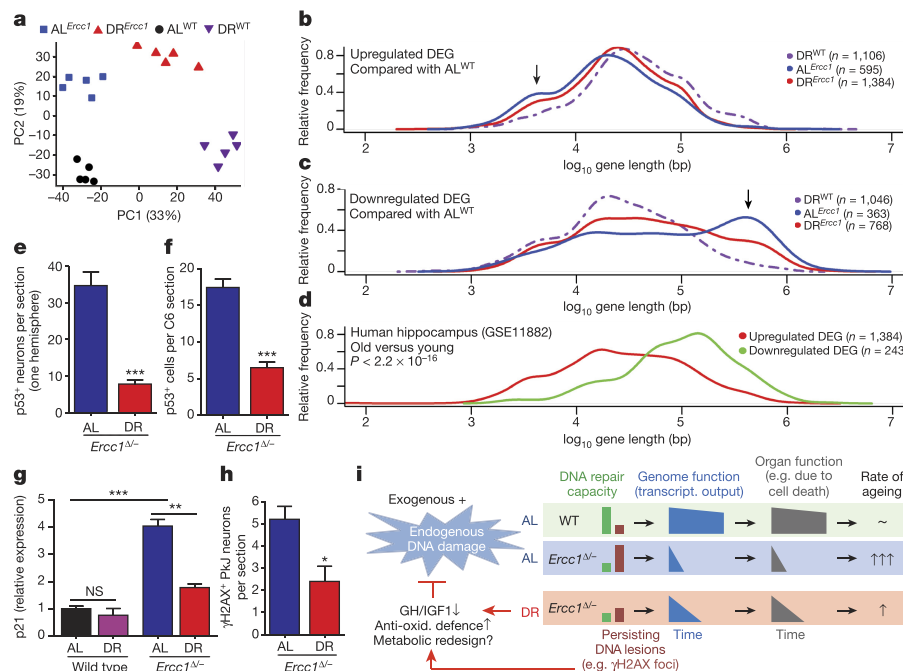


Figure 3 | Dietary restriction preserves genome function. **a**, Principal component analysis (PCA) of full genome liver RNA expression profiles of 11-week old *AL*^{*Ercc1*} mice (blue squares), *DR*^{*Ercc1*} (red triangles), *AL*^{WT} (black circles) and *DR*^{WT} (purple triangles) mice. This analysis takes into account all the genes in the microarray platform. The two main principal components, PC1 and PC2, explain 52% of the variability in the original data set: PC1 (*x* axis, 33%) differentiates on the basis of expression changes induced by dietary restriction, independent of genotype; PC2 (*y* axis, 19%) reflects differences associated with genotype. **b, c**, Relative frequency plot of gene length (log scale) in *DR*^{WT} versus *AL*^{WT} (purple dashed lines), *AL*^{*Ercc1*} versus *AL*^{WT} (blue lines) and *DR*^{*Ercc1*} versus *AL*^{WT} (red lines). **b**, Only upregulated genes. **c**, Only downregulated genes. Black arrows, extra peak of upregulated short genes (**b**) and peak of downregulated long genes (**c**) in *AL*^{*Ercc1*} mice. **d**, Relative frequency plot of gene length of DEGs in hippocampus from ~80-year-old humans versus ~20-year-old humans. Red, upregulated genes; green, downregulated genes. The DEGs

from human hippocampus were selected using a log₂-fold change cut-off of 0.5 and FDR < 0.05. The data set used corresponds to NCBI gene expression omnibus, number GSE11882. **e, f**, p53-positive cells counted in the neocortex of three consecutive transverse brain sections (**e**) at the level of the bregma (Mouse Brain Atlas, Paxinos) and three consecutive C6 cervical spinal cord sections (**f**). Sections from 16-week-old *DR*^{*Ercc1*} mice (*n* = 4) show significantly reduced levels of p53-positive cells (*P* < 0.0001 for neocortex and *P* = 0.0002 for spinal cord) than sections from *AL*^{*Ercc1*} mice (*n* = 4). **g**, Relative expression changes in the p53 target gene *p21* in 11-week-old wild-type and *Ercc1*^{Δ/Δ} mice induced by dietary restriction (*n* = 5). **h**, γH2AX-positive Purkinje (Pk) neurons were counted in cerebellum of five consecutive transverse brain sections from *AL*^{*Ercc1*} and *DR*^{*Ercc1*} mice (*P* = 0.014). Error bars indicate mean ± s.e. **P* < 0.05, ***P* < 0.01, ****P* < 0.001. **i**, Mechanistic model for the anti-ageing effect of dietary restriction.

Table 1 | Number of DEGs under dietary restriction

Comparison	DEG	Up	Down	Ratio up:down	Concordant	Discordant
DR ^{WT} vs. AL ^{WT}	2,704	1,522	1,182	1.29*		
DR ^{Ercc1} vs. AL ^{Ercc1}	1,106	669	437	1.53*		
Common	688	391	293		684	4
Enrichment factor	10.3	17.3	25.6			

The number of differentially expressed gene (DEG) probes are shown for wild-type (WT) and *Ercc1*^{Δ/−} liver under dietary restricted (DR) and *ad libitum* (AL) conditions. **P* = 0.019 (Chi-square test with Yates correction).

consistent with overlapping Gene Ontology pathways and transcription factors (Supplementary Table 2). Molecular analyses of insulin, mTOR and GH/IGF1 signalling pathways and microRNA expression, amongst others, further support the parallels between DR^{WT} and DR^{Ercc1} gene expression (for example, the *GHR* gene, already suppressed in AL^{Ercc1} mice is further reduced upon dietary restriction; Extended Data Figs 1f, 7, 8a–e). These data revealed also an unexpected link between *Ercc1*-related DNA-repair deficiency and the unfolded-protein response (Extended Data Table 1 and Supplementary Table 2).

Transcription-coupled repair (TCR) deficiencies, which prevent resumption of RNA synthesis after transcription-blocking DNA damage in *Ercc1*^{Δ/−} mice, other mouse mutants and human patients, appear to be uniformly associated with premature ageing, affecting non- or slow-proliferating organs such as the nervous system, liver and kidney in particular². This suggests that time-dependent accumulation of transcription-blocking lesions contribute to accelerated ageing. Since DNA damage occurs stochastically, long genes generally accumulate more lesions and consequently become more transcriptionally crippled than short genes. Indeed, the DEGs of AL^{Ercc1} liver gene expression profiles at 11 weeks displayed a highly significant bias for long genes to be overrepresented in the class of downregulated genes and underrepresented in the category of upregulated genes (Fig. 3b, c, arrows). These and other data (Table 1 and Extended Data Table 2) suggest genome-wide accumulation of transcription-stalling lesions. We observed a similar, albeit milder, expression bias upon normal ageing in rat liver and human hippocampus (Fig. 3d and Extended Data Fig. 8f). Notably, dietary restriction in *Ercc1*^{Δ/−} mice strongly retarded this expression shift (red curve in Fig. 3b, c, and Extended Data Table 2).

That there were reduced DNA damage loads in DR^{Ercc1} mice is supported by an observed reduction in p53-expressing cells in the brain (Fig. 3e, f and Extended Data Fig. 6d), diminished apoptosis of retinal photoreceptors (Extended Data Fig. 6a) and preservation of neurons (Fig. 2f, g). We also found evidence for the suppression of p53 transcriptional activity in the liver of DR^{Ercc1} mice (Supplementary Table 2), further supported by dietary restriction-induced downregulation of expression of the microRNA miR-34a, a target of p53 (Extended Data Fig. 8c). Additionally, key senescence parameters, which were elevated in AL^{Ercc1} mice, were mitigated by dietary restriction. These include *p21* (also known as *Cdkn1a*) (Fig. 3g), *p16* (*Cdkn2a*) and *Il6* (Extended Data Fig. 8e). The proportion of Purkinje cell nuclei containing γH2AX foci, which reflect DNA breaks, appeared significantly reduced (Fig. 3h and Extended Data Fig. 6e). We conclude that dietary restriction concomitantly delays ageing, attenuates accumulation of genome-wide DNA damage and preserves transcriptional output, in all probability contributing to improved cell viability in *Ercc1* mutants (Figs 2f, g and 3c, h).

It seems unlikely that dietary restriction can overcome defects in DNA repair. It may be that dietary restriction reduces the induction and/or alters responses to damage, to which DNA repair mutants may over-respond. Indeed, the expression profiles of diet-restricted mice (Extended Data Figs 7, 8 and Supplementary Table 2) suggest that dietary restriction increases resistance to DNA damage-induced stress, improves antioxidant defences, alters insulin and other hormonal signalling pathways (redesigning major metabolic routes such

as glycolysis, oxidative phosphorylation and the pentose phosphate pathway), alters mitochondrial function and apoptotic responses and induces a shift from pro- to anti-inflammatory cytokines^{10,24}. Figure 3i presents a model that integrates our findings on dietary restriction, DNA damage and (accelerated) ageing. DNA damage from exogenous and endogenous sources accumulates with ageing and is accelerated in repair-deficient *Ercc1*^{Δ/−} mice. Stochastic DNA lesions reduce transcriptional output in a gene-size-dependent manner leading to cell dysfunction and death, stem cell exhaustion, organ and tissue atrophy and functional decline, which together cause ageing-related diseases. Both accumulation of DNA damage and dietary restriction trigger an anti-ageing response, which involves suppression of growth, upregulation of anti-oxidant defences and, presumably, metabolic redesign. This response reduces steady-state levels of reactive metabolites, thereby preserving genome integrity and delaying ageing-related functional decline.

Post-mitotic neurons have to reconcile one of the highest metabolic rates of the body with the preservation of a delicate homeostatic balance for over a century in humans. While everything else in cells can be turned over, the up-to-10⁵ daily DNA lesions per cell^{25,26} can only be repaired, requiring continuous efficient repair. That TCR defects are linked with severe neurodegeneration in Cockayne syndrome and trichothiodystrophy (TTD) indicates that removing transcription-obstructing lesions from the genome is vital for neuronal survival²³. Using tissue-specific repair mutants, we have shown that neurodegeneration is at least partly cell-autonomous^{6,27,28}, consistent with endogenous DNA damage being a main driver. The strong protection given by dietary restriction indicates that neurons possess considerable reserves to restrict DNA damage and prevent cell death. Cell-intrinsic mechanisms and systemic inflammatory and hormonal responses modulated by dietary restriction may contribute to the remarkable resilience of the neuronal system; they protect 50% more neurons from death, postpone the onset of tremors, imbalance and paresis and fully preserve motor performance (Fig. 2 and Supplementary Videos 1, 2). The notable preservation of neuronal function by dietary restriction in the *Ercc1*^{Δ/−} mutant is consistent with increasing evidence for the beneficial effects of dietary restriction in animal models of various neurodegenerative disorders^{29,30} and opens the door to nutritional and pharmacological interventions that could prevent the onset of these devastating diseases.

The strong evolutionary conservation of the dietary-restriction response, and the notable parallels between mouse and human symptoms, mean it is likely that the effect of dietary restriction is preserved from mouse to man. Therefore, an obvious future application may be the counterintuitive use of dietary restriction or pharmaceutical mimetics to treat DNA-repair-defective progeroid syndromes. Besides Cockayne syndrome and TTD, this may apply to xeroderma pigmentosum, combined xeroderma pigmentosum/Cockayne syndrome, cerebro-oculo-facio-skeletal (COFS) syndrome, Fanconi's anaemia, the RecQ-helicase-dysfunction driven family of conditions (Werner, Bloom's and Rothmund Thomson syndrome), ataxia telangiectasia, and others. Additionally, repair-deficient mice could prove to be useful tools for understanding anti-ageing interventions and helping to find alternatives to dietary restriction by strongly reducing the labour, time, costs and number of animals required, and also for defining the long-term health effects of nutrition.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 June 2015; accepted 25 July 2016.

Published online 24 August 2016.

1. Niedernhofer, L. J. *et al.* A new progeroid syndrome reveals that genotoxic stress suppresses the somatotrophic axis. *Nature* **444**, 1038–1043 (2006).
2. Marteijn, J. A., Lans, H., Vermeulen, W. & Hoeijmakers, J. H. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat. Rev. Mol. Cell Biol.* **15**, 465–481 (2014).

3. Dollé, M. E. *et al.* Broad segmental progeroid changes in short-lived *Erc1*(Δ 7) mice. *Pathobiol. Aging Age Relat. Dis.* **1**, 10.3402/pba.v1i0.7219 (2011).
4. Vermeij, W. P., Hoeijmakers, J. H. & Pothof, J. Genome integrity in aging: human syndromes, mouse models, and therapeutic options. *Annu. Rev. Pharmacol. Toxicol.* **56**, 427–445 (2016).
5. Garinis, G. A. *et al.* Persistent transcription-blocking DNA lesions trigger somatic growth attenuation associated with longevity. *Nat. Cell Biol.* **11**, 604–615 (2009).
6. Barnhoorn, S. *et al.* Cell-autonomous progeroid changes in conditional mouse models for repair endonuclease XPG deficiency. *PLoS Genet.* **10**, e1004686 (2014).
7. Colman, R. J. *et al.* Caloric restriction delays disease onset and mortality in rhesus monkeys. *Science* **325**, 201–204 (2009).
8. Fontana, L., Partridge, L. & Longo, V. D. Extending healthy life span—from yeast to humans. *Science* **328**, 321–326 (2010).
9. Barzilai, N., Huffman, D. M., Muzumdar, R. H. & Bartke, A. The critical role of metabolic pathways in aging. *Diabetes* **61**, 1315–1322 (2012).
10. Speakman, J. R. & Mitchell, S. E. Caloric restriction. *Mol. Aspects Med.* **32**, 159–221 (2011).
11. Hoeijmakers, J. H. DNA damage, aging, and cancer. *N. Engl. J. Med.* **361**, 1475–1485 (2009).
12. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194–1217 (2013).
13. van der Pluijm, I. *et al.* Impaired genome maintenance suppresses the growth hormone–insulin-like growth factor 1 axis in mice with Cockayne syndrome. *PLoS Biol.* **5**, e2 (2007).
14. Wijnhoven, S. W. *et al.* Accelerated aging pathology in ad libitum fed *Xpd*(TTD) mice is accompanied by features suggestive of caloric restriction. *DNA Repair (Amst.)* **4**, 1314–1324 (2005).
15. Schumacher, B. *et al.* Delayed and accelerated aging share common longevity assurance mechanisms. *PLoS Genet.* **4**, e1000161 (2008).
16. Weeda, G. *et al.* Disruption of mouse *ERCC1* results in a novel repair syndrome with growth failure, nuclear abnormalities and senescence. *Curr. Biol.* **7**, 427–439 (1997).
17. de Waard, M. C. *et al.* Age-related motor neuron degeneration in DNA repair-deficient *Erc1* mice. *Acta Neuropathol.* **120**, 461–475 (2010).
18. Partridge, L. & Gems, D. Benchmarks for ageing studies. *Nature* **450**, 165–167 (2007).
19. Mair, W., Goymer, P., Pletcher, S. D. & Partridge, L. Demography of dietary restriction and death in *Drosophila*. *Science* **301**, 1731–1733 (2003).
20. Bogliolo, M. *et al.* Mutations in *ERCC4*, encoding the DNA-repair endonuclease XPF, cause Fanconi anemia. *Am. J. Hum. Genet.* **92**, 800–806 (2013).
21. Kashiwama, K. *et al.* Malfunction of nuclease *ERCC1-XPF* results in diverse clinical manifestations and causes Cockayne syndrome, xeroderma pigmentosum, and Fanconi anemia. *Am. J. Hum. Genet.* **92**, 807–819 (2013).
22. Jaarsma, D., van der Pluijm, I., van der Horst, G. T. & Hoeijmakers, J. H. Cockayne syndrome pathogenesis: lessons from mouse models. *Mech. Ageing Dev.* **134**, 180–195 (2013).
23. Wilson, B. T. *et al.* The Cockayne Syndrome Natural History (CoSyNH) study: clinical findings in 102 individuals and recommendations for care. *Genet. Med.* **18**, 483–493 (2016).
24. Bartke, A., Sun, L. Y. & Longo, V. Somatotrophic signaling: trade-offs between growth, reproductive development, and longevity. *Physiol. Rev.* **93**, 571–598 (2013).
25. Gates, K. S. An overview of chemical processes that damage cellular DNA: spontaneous hydrolysis, alkylation, and reactions with radicals. *Chem. Res. Toxicol.* **22**, 1747–1760 (2009).
26. Swenberg, J. A. *et al.* Endogenous versus exogenous DNA adducts: their role in carcinogenesis, epidemiology, and risk assessment. *Toxicol. Sci.* **120** (Suppl. 1), S130–S145 (2011).
27. Borgesius, N. Z. *et al.* Accelerated age-related cognitive decline and neurodegeneration, caused by deficient DNA repair. *J. Neurosci.* **31**, 12543–12553 (2011).
28. Raj, D. D. *et al.* Priming of microglia in a DNA-repair deficient model of accelerated aging. *Neurobiol. Aging* **35**, 2147–2160 (2014).
29. Stranahan, A. M. & Mattson, M. P. Metabolic reserve as a determinant of cognitive aging. *J. Alzheimers Dis.* **30** (Suppl. 2), S5–S13 (2012).
30. Halagappa, V. K. *et al.* Intermittent fasting and caloric restriction ameliorate age-related behavioral deficits in the triple-transgenic mouse model of Alzheimer's disease. *Neurobiol. Dis.* **26**, 212–220 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank P. de With, J. Rigters, E. Haasdijk, S. Gabriels, E. J. M. Stynenbosch, N. van Vliet, Y. van Loon, J. Baan and the animal caretakers for general assistance with mouse experiments. We thank A. H. J. Danser and J. P. van Leeuwen for support. We acknowledge financial support from the National Institute of Health (NIH)/National Institute of Ageing (NIA) (1P01 AG-17242-02), the National Institute for Public Health and the Environment and the Ministry of Health, Welfare and Sport of The Netherlands (S/340005), European Research Council Advanced Grant DamAge and Proof of Concept Grant Dementia to J.H.J.H., the European commission FP7 Markage (FP7-Health-2008-200880), DNA Repair (LSHG-CT-2005-512113), EU ITN Address (GA-316390), the KWO Dutch Cancer Society (5030), the Dutch CAA Foundation and the Royal Academy of Arts and Sciences of the Netherlands (academia professorship to J.H.J.H.). The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement number HEALTH-F2-2010-259893. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions W.P.V., M.E.T.D., E.R., J.V., H.v.S., and J.H.J.H. designed the research and wrote the manuscript. D.J., C.P.-G., A.J.M.R., S.M.B., B.C.v.d.E., A.d.B., Á.G., and J.P. contributed to editing the manuscript. W.P.V., M.E.T.D., E.R., R.M.C.B., and S.B. performed and analysed the mouse lifespan cohorts. E.R., B.N., C.T.v.O., R.M.C.B., S.B., and S.I. performed genotyping and coordinated animal sectioning. S.A.Y., R.V.K., and A.d.B. assessed the ageing pathology characteristics. R.V.K. and S.I. performed FACS analysis of nuclei. S.M.B. and B.C.E. quantified bone changes. H.W. and A.J.M.R. quantified vascular function. C.R.B. performed the immunological analyses. W.P.V., D.J., R.M.C.B., and S.B. performed and analysed phenotypical scoring and behavioural analysis. W.P.V. and D.J. characterized neuropathological changes. W.P.V., E.R., C.P.-G., C.T.v.O., J.L.A.P., Á.G., and J.P. performed transcriptomic analyses and analysed the data. W.P.V., M.E.T.D., E.R., C.T.v.O., R.M.C.B., S.B., and S.I. performed the molecular studies.

Author Information The expression data have been deposited to the Gene Expression Omnibus database under accession number GSE77495. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.E.T.D. (martijn.dolle@rivm.nl) or J.H.J.H. (j.hoeijmakers@erasmusmc.nl).

Reviewer Information *Nature* thanks C. Lopez-Otin, M. Mattson and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Mouse models. The generation and characterization of *Ercc1*^{Δ/+} and *Ercc1*^{Δ/-} mice have been previously described¹⁶. *Ercc1*^{Δ/-} mice were obtained by crossing *Ercc1*^{Δ/+} (in a pure C57BL6J or FVB background) with *Ercc1*^{+/-} mice (in a pure FVB or C57BL6J background respectively) to yield *Ercc1*^{Δ/-} offspring with a genetically uniform F1 C57BL6J/FVB hybrid background (see ref. 6 for motivation). Wild-type F1 littermates were used as controls. *Xpg*^{-/-} mice have been characterized previously⁶ and were generated by crossing *Xpg*^{+/-} (in a pure C57BL6J background) with *Xpg*^{+/-} mice (in a pure FVB background). Hence, all animals used in the studies described here were of the same F1 C57BL6J/FVB hybrid background. Typical unfavourable characteristics, such as blindness in an FVB background or deafness in a C57BL6J background, do not occur in this hybrid background.

Mice were weighed, visually inspected weekly, and scored in a blinded fashion for gross morphological and motor abnormalities. Since the *Ercc1*^{Δ/-} and *Xpg*^{-/-} mice were smaller, food was administered within the cages and water bottles with long nozzles were used from around two weeks of age. Animals were maintained in a controlled environment (20–22 °C, 12 h light:12 h dark cycle) and were housed in individual ventilated cages under specific pathogen free conditions. Animals were individually housed at the EMC location and group housed at the RIVM location. Experiments were performed in accordance with the Principles of Laboratory Animal Care and with the guidelines approved by the Dutch Ethical Committee in full accordance with European legislation.

For the lifespan studies the indicated number of mice per group for *ad libitum* and 30% dietary restriction were generated. Additionally, several cross-sectional cohorts were generated. For *Ercc1*^{Δ/-} mice we generated groups which were killed at 7, 11, 16 or 30 weeks of age. The 7-week group consisted only of *ad libitum*-fed animals while the 30-week group consisted only of dietary restriction-treated mice. For wild-type mice, *ad libitum*-fed and dietary restriction-treated groups were sacrificed at 11, 16 or 20 weeks. Sample size of the lifespan cohorts were based on power analysis. No statistical methods were used to predetermine sample size of cross-sectional cohorts. Animals were divided randomly over all groups to prevent selection bias. All mice were clinically diagnosed daily in a blinded manner and, when moribund, killed, after which necropsy was performed. Animals from cross-sectional cohorts were killed when necropsy age was reached. Organs were stored at -80 °C for molecular analysis or (perfusion) fixated in (para) formaldehyde for pathological examinations. Statistics was performed with survival curve analysis using the product-limit method of Kaplan and Meier in GraphPad Prism.

Diets. All animals were bred and maintained on AIN93G synthetic pellets (Research Diet Services B.V.; gross energy content 4.9 kcal/g dry mass, digestible energy 3.97 kcal/g). The initial lifespan cohort, shown in Fig. 1a, were fed standard AIN93G pellets containing 2.5 g/kg choline bitartrate. To avoid potential formation of bladder and kidney stones, we replaced choline bitartrate with choline chloride in all subsequent experiments. The amount of dietary restriction was determined in a prior pilot study and food intake of the *ad libitum*-fed mice was continuously monitored. On average, *Ercc1*^{Δ/-} and *Xpg*^{-/-} mice ate 2.3 g food per day. Dietary restriction was initiated at 7 weeks of age with 10% food reduction (2.1 g/day), when animals reached almost-maximum bodyweight and development was completed. Dietary restriction was increased weekly by 10%, until it reached 30% dietary restriction (1.6 g/day) from 9 weeks of age onward. Temporary dietary restriction was initiated directly with 30% food reduction at 6 weeks of age. These mice received *ad libitum* food again from 12 weeks onward. Wild-type mice ate on average 3.0 g food per day, resulting in 2.1 g/day for 30% dietary restriction. Food was given to the animals just before the start of the dark (active) period to avoid alteration of the biological clock.

Pathology assessment of ageing characteristics. Representative sections from the liver, kidneys, sciatic nerve, testes and femur were processed, stained with haematoxylin and eosin, and microscopically examined in a blinded manner by two board-certified pathologists (SAY, AdB) for the presence of histopathologic lesions. The severity score of lesions was semi-quantitatively assessed. Scores were given as absent (0), subtle (1), mild (2), moderate (3), severe (4), and massive (5). Digital images from the kidneys and femur cortical bone at mid-shaft area were taken for morphometric analysis using Labsense image analysis software (Olympus). Ageing characteristics were assessed in >5 animals per group per sex. Groups were compared with nonparametric Mann–Whitney U and Kruskal–Wallis tests.

FACS analysis of nuclear DNA content. Polyploidy levels were assessed based on propidium iodide (PI) fluorescence using FACS analysis^{31,32}. A small part of the left lobe (approximately 5 mm³) was dissected from *ad libitum*- and dietary restriction-treated *Ercc1*^{Δ/-} mice (7, 11, 16 and 30 weeks, *n* = 5) and wild-type mice (11 weeks, *n* = 5), cut into small fragments and suspended in 800 μl PBS

using a syringe (21G). 300 μl homogenate was added to 300 μl 100% ethanol for fixation. Samples were stored for at least 24 h before further processing. After fixation the liver homogenate was washed with ice-cold PBS and subsequently incubated with a pepsin solution for 20 min. After washing in PBS/Tween-20, cells were collected in 500 μl PBS supplemented with 5 μg/ml PI and 250 μg/ml RNase and samples were measured using the FACS (FACSCalibur, Becton Dickinson). Differences between groups were assessed with a two-way ANOVA, with age and diet as fixed factors.

Micro-computed tomographic (micro-CT) quantification of bone thickness. *Ad libitum*- and diet-restricted mice were killed by cervical dislocation at scheduled ages, femora were excised and non-osseous tissue was removed. Two days after fixation in 4% formalin, the right femora were scanned using Skyscan 1076 *in vivo* X-Ray computed tomography (Bruker microCT) with a voxel size of 8.88 μm. Osseous tissue was distinguished from non-osseous tissue by segmenting the reconstructed grayscale images with an automated algorithm using local thresholds³³. The region of interest (ROI) (the distal metaphysis of the femora) was selected using 3D data analysis software. To compensate for bone length differences, the length of each ROI was determined relative to the largest specimen femur of the cohort. The cortex and trabeculae of the metaphysis were separated using automated software developed in-house. The thickness of the trabeculae and cortices were assessed using 3D analysis software as described³⁴ using the CT analyser software package (Bruker microCT). A bone specimen with known bone morphometrics was included within each scan as a quantitative control. Statistical significance was calculated using one-way Anova with Bonferroni's multiple comparison test.

Ex vivo vascular function. The responses of isolated aortic tissue were *ex vivo* measured in small-wire myograph organ baths containing oxygenated Krebs-Henseleit buffer at 37 °C. After precontraction with 30 nmol/l U46619, relaxation concentration–response curves to acetylcholine were constructed³⁵.

Immunological analyses. Single-cell suspensions were prepared from spleen by passing the cells through a cell strainer with HEPES-buffered saline solution (HBSS) supplemented with 2% FBS and washed. Erythrocytes were eliminated with ACK buffer. For CD4⁺CD25⁺Foxp3⁺ staining, cells were first stained for the expression of cell surface markers and then fixed, permeabilised, and stained using the Foxp3 kit (eBiosciences) according to the manufacturer's instructions. FACS analysis was performed using FACS (Becton Dickinson) and analysed with FlowJo Software (TreeStar).

Blood glucose, insulin and albumin levels. Mice were killed by CO₂ asphyxiation and blood was immediately collected from the heart. Glucose levels were measured using a Freestyle mini blood glucose metre. Insulin and albumin levels were measured in blood plasma using an ultrasensitive mouse insulin Elisa (Mercodia AB) or mouse albumin ELISA kit (Immunology Consultants Laboratory, Inc.), respectively. Insulin levels were determined after overnight fasting. Glucose levels were determined after feeding, at the beginning of the dark period.

IgA measurements. Euthanasia of moribund or cross-sectional animals was performed by intramuscular injection of a ketamine–rompun mixture, followed by exsanguination³. IgA immunoglobulin was measured in blood serum using the commercially available bead-based multiplexed panel Mouse Immunoglobulin Isotyping (Millipore Corporation). Standard analysis protocols were followed and all samples were analysed at least in *duplo*.

Phenotype scoring. The mice were weighed and visually inspected weekly, and were scored in a blinded manner by two experienced research technicians (R.M.C.B. and S.B.) for the onset of various phenotypical parameters. Clasping was measured by suspending mice by their tails for 20 s. A clasping event was scored when retraction of both hind limbs towards the body was observed for at least 5 s. Whole-body tremor was scored if mice were trembling for a combined total of at least 10 s when put on a flat surface for 20 s. Impaired balance was determined by observing the mice walking on a flat surface for 20 s. Mice that had difficulties in maintaining an upright orientation during this period were scored as having imbalance. If mice showed a partial loss of function of the hind limbs, they were scored as having paresis. Statistics were performed with survival-curve analysis using the product-limit method of Kaplan and Meier in GraphPad Prism.

Behavioural analyses. Rotarod performance was assessed by measuring the average time spent on an accelerating rotarod (Ugo Basile). All animals were given four consecutive trials of a maximum of 5 min with inter-trial intervals of 1 h. For weekly monitoring, the motor coordination performance was measured with two consecutive trials of a maximum of 5 min. Grip strength was determined by placing mice with forelimbs or all limbs on a grid attached to a force gauge, and steadily pulling the mice by their tail. Grip strength is defined as the maximum strength produced by the mouse before releasing the grid. For each value the test was performed in triplicate.

TUNEL staining. To quantify apoptotic cells in the retina, eyes were fixed overnight in 10% phosphate-buffered formalin (JT Baker), paraffin-embedded, sectioned at 5 mm, and mounted on Superfrost Plus slides. Paraffin sections were employed for TdT-mediated dUTP nick-end labelling (TUNEL) assay using an Apoptag Plus Peroxidase *in situ* apoptosis detection kit (Millipore). Sections were deparaffinised and incubated as described by the manufacturer. Statistical differences were calculated with a *t*-test.

Antibodies. Primary antibodies (supplier; catalogue number; dilutions) used in this study were as follows: rabbit anti-ATF3 (Santa Cruz; sc-188; 1:2,000), goat anti-ChAT (Millipore; AB144P; 1:500); rabbit anti-GFAP (DAKO; Z0334; 1:8,000); mouse anti-GM130 (BD Transduction; 610823; 1:100); rabbit anti-Iba-1 (Wako; 019-19741; 1:5,000); rat anti-Mac2 (Cedarlane; CL8942AP; 1:2,000); mouse anti-NeuN (Millipore; MAB377; 1:1,000); rabbit anti-p53 (Leica; NCL-p53-CM5p; 1:1,000); mouse anti- γ H2AX (Millipore; 05-636; 1:4,000). For avidin-biotin-peroxidase immunocytochemistry biotinylated secondary antibodies from Vector Laboratories, diluted 1:200 were used. Alexa488-, Cy3-, and Cy5-conjugated secondary antibodies raised in donkey (Jackson ImmunoResearch) diluted at 1:200 were used for confocal immunofluorescence.

Histological procedures. Mice were anaesthetized with pentobarbital and perfused transcardially with 4% paraformaldehyde. The brain and spinal cord were carefully dissected out, post-fixed for 1 h in 4% paraformaldehyde, cryoprotected, embedded in 12% gelatin, rapidly frozen, and sectioned at 40 μ m using a freezing microtome or stored at -80°C until use. Frozen sections were processed free floating using the ABC method (ABC, Vector Laboratories) or single-, double-, and triple-labelling immunofluorescence. Immunoperoxidase-stained sections were analysed and photographed using an Olympus BX40 microscope. Immunofluorescence sections were analysed using a Zeiss LSM700 confocal microscope. Mean intensities were quantified using Fiji. Statistical differences were calculated with a *t*-test.

RNA isolation. Total RNA was extracted using QIAzol lysis Reagent from mouse tissue specimens. For increased purity, miRNAeasy Mini Kits (QIAGEN) were used. Addition of wash buffers RPE and RWT (QIAGEN) was done mechanically by using the QIAcube (QIAGEN) via the miRNeasy program and tissue was stored at -80°C . The concentration of RNA was measured by Nanodrop (Thermo Fisher Scientific).

Real-time PCR. Gene expression analyses were performed with gene-specific real-time PCR primers (see below) using SYBR Green (Sigma-Aldrich) and Platinum Taq polymerase (Life Technologies) on a Bio-Rad CFX96 thermocycler or with pre-designed TaqMan Gene Expression Assays (given below) with a 7500 Fast Real-Time PCR System (Applied Biosystems). Relative gene expressions were calculated as previously described⁶. For SYBR Green method the following primers were used (forward primer 5' to 3'; reverse primer 5' to 3'): Gsta1 (CTTCTGACCCCTTCCCTCT; ATCCATGGGAGGCTTCTCT), Nqo1 (GGTAGCGGTCCATGTACT; GAGTGTGGCCAATGCTGTAA), Nfe2l2 (AGGACATGGAGCAAGTTTGG; TCTGTCACTGTGGCTTCTGG), Gsta2 (CGAGCAATTCTCCAGGTGA; TATTCGTGGACTTGGGCACG), Fkbp5 (TGTTCAAGAAGTTCGACAGAGC; CCTTCTTGCTCCCAGCTTT), Srxn1 (TGAGCAGCTCCTCTGATGTG; GCTGAGGTGACAATTGACTATGG), Gsta4 (TCGATGGGATGATGCTAC; CATCTGCATACATGTAACCTCG), Gdln (TGGAGCAGCTGTATCAGTGG; CAAAGGCAGTCAAATCTGGTG), Hmxo1 (CAGGTGATGCTGACAGAGGA; ATGGCATAAATCCCACTGC), Gclc (AGATAGTAGAACACGGGAGGAG; TGATCCTAAAGCGATTGTCTTC), Ephx1 (GATGGGAGGAAGTGCACACC; AGCAGAGAAGCCAGGATGA), Mgst1 (CTCGGCAGGACAACCTTGC; CCATGCTTCCAATCTTGGTC), Tubg2 (CAGACCAACCACTGCTACAT; AGGGAATGAAGTTGGCCAGT), Hprt (TGATAGATCCATTCCTATGACTGTAGA; AAGACATTCTTCCAGTTAAAGTTGAG), Rps9 (ATCCGCCAACGTACATTA; TCTTCACTCGGCCTGAC). As pre-designed TaqMan assays we used (order number; sequence 5' to 3'): Ghr (Mm00439093_m1; GACAAGCTGCAAGAATTGCTCATGA), Igflr (Mm00802831_m1; GGCCAGAAGTGAGCAGAATAATCT), HPRT-E2_3 (HPRT-E2_3_F; GCCGAGGATTGGAAAAAGTGTTTA, HPRT-E2_3_R; TTCATGACATCTCGAGCAAGTCTTT, HPRT-E2_3_M; CAGTCTGTCCAATATCA), POLR2A-E2_3 (POLR2A-E2_3F; GCAGTTCGGAGTCTCTGAGT, POLR2A-E2_3R; CCTCTGTGTTTCTGGGTATTTGA, POLR2A-E2_3M2; CATCCGCTTCAATTCAT).

Microarray hybridizations. RNA quality was assessed using the 2100 Bio-Analyzer (Agilent Technologies) following the manufacturer's instructions. The quality of the RNA is expressed as the RNA integrity number (RIN, range 0–10). Samples with a RIN below 8 were excluded from analysis. Hybridization to Affymetrix HT MG-430 p.m. Array Plates was performed at the Microarray Department of the University of Amsterdam according to Affymetrix protocols. Quality control and normalization were performed using the pipeline at the www.arrayanalysis.org website (Maastricht University).

miRNA expression. The same total RNA extracts were used as extracted for mRNA analysis (above). miRNA expression levels were assessed using a miRNA micro-array (miRCURY LNA microRNA Array (7th Gen.), Exiqon). All probes with more than three calls were selected for assessing differential expression between groups. Differences in mean expression were compared using a one-way ANOVA. Probes with a FDR of 5% were considered as significantly differentially expressed.

Total RNA-seq. RNA expression analysis was also performed with the next-generation sequencing approach on one animal per treatment as described in ref. 36.

Data pre-analysis. Raw data (CEL files) were normalized by robust multichip average (RMA) in the oligo BioConductor package, which summarizes perfect matches through median polish and collapses probes into core transcripts based on CDF annotation file provided by Affymetrix using the R open statistical package (<http://www.r-project.org/>). All data files have been submitted to the NCBI gene expression omnibus under accession number GSE77495.

Principal component analysis. Principal component analysis (PCA) was performed using all the probe sets in the array. A graphical representation was generated to show the relationship among the different samples. PCA is a linear projection method that defines a new dimensional space to capture the maximum information present in the initial data set. It is an unsupervised exploratory technique used to remove noise, reduce dimensionality and identify common/dominant signals oriented to try to find biological meaning³⁷. The two principal components with the highest amount of variance were plotted. PCA was performed using the prcomp package and the plot was drawn with gplots, both from the Bioconductor project (<https://www.bioconductor.org/>).

Detection of differentially expressed genes (DEG). The linear model from Limma³⁸ implemented in R was used to identify the DEGs. Pairwise comparisons for each genotype between ad libitum and dietary restriction samples were applied to calculate the fold change (FC), *P* value and false discovery rate (FDR) for each probe in the microarray. Cut-off values for a DEG were put at FDR < 5% with FC \geq |1.5|. For all mouse analyses, differentially expressed probes were considered as DEGs.

Determination of enrichment factor and *P* values of overlap. Overlap between lists of DEGs was identified looking by the intersection between pair of lists. To determine if the overlap was higher than expected by chance the hypergeometric distribution was used as is implemented in phyper function in R. Additionally the factor of enrichment was calculated with the formula: $EF = nAB / ((nA \times nB) / nC)$. Where: nA = Number of DEG in experimental group A; nB = Number of DEG in experimental group B; nC = Number of total genes in the microarray; nAB = Number of common DEG between A and B.

Pathway analysis. Pathway enrichment analysis was conducted via overrepresentation analysis (ORA). ORA was performed in the Interactive pathway analysis (IPA) of complex genomics data software (Ingenuity Systems, Qiagen) by employing a pre-filtered list of differentially expressed genes. Genes were selected as differentially expressed if they had a fold change \geq 1.5 and an FDR lower than 0.05. The over-represented canonical pathways were generated based on information in the Ingenuity Pathways Knowledge Base. A pathway was selected as deregulated when the *P* value in the Fisher test was lower than 0.01.

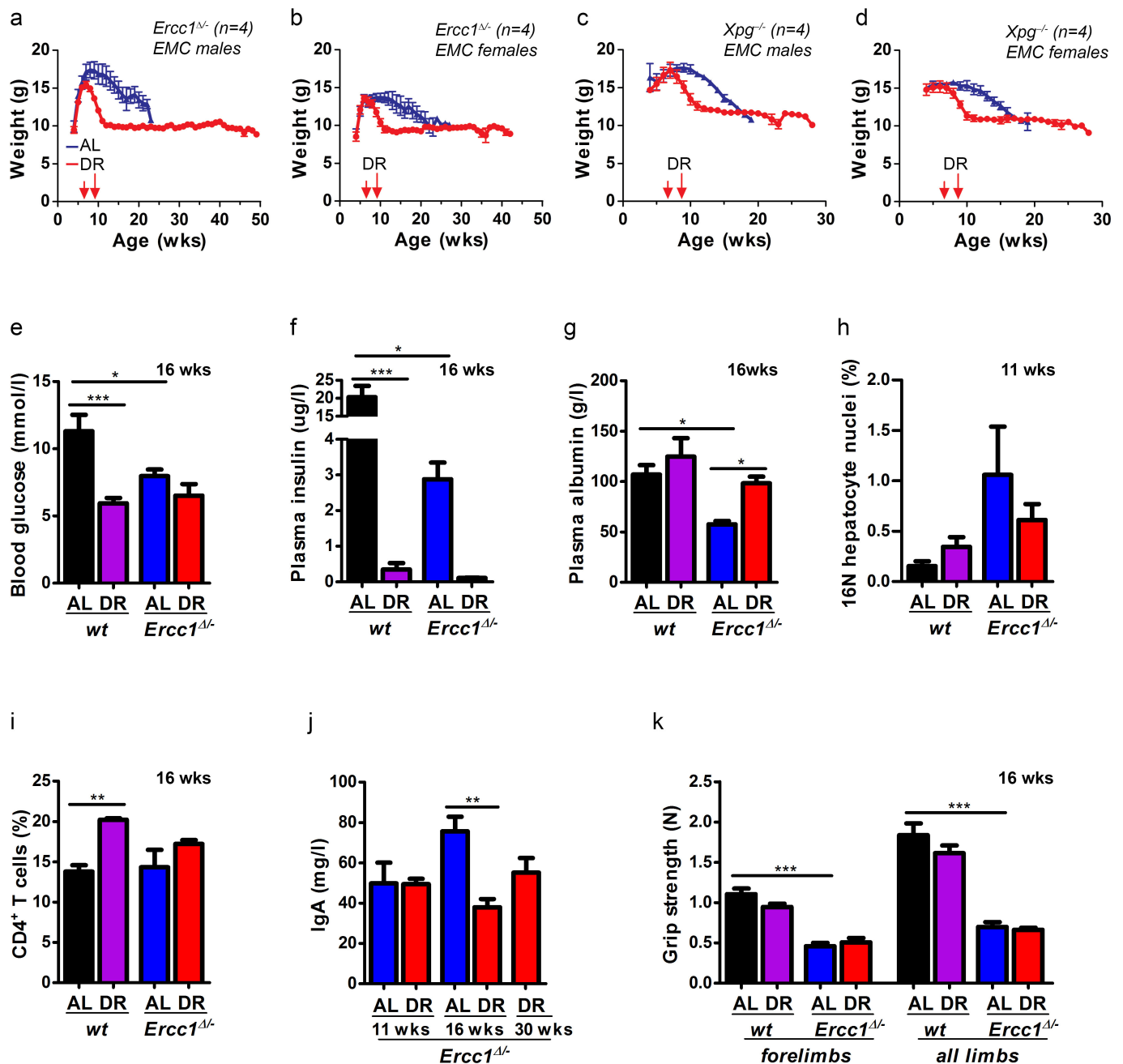
Additionally, IPA transcription factor (TF) analysis was performed to identify the cascade of upstream transcriptional regulators that can explain the observed gene expression changes in the different lists of DEGs. To do this, data stored in the Ingenuity Knowledge Base, with prior information on the expected effects between TF and their target genes, were used. The analysis examines how many known targets of each TF are present in the list of DEGs, and also compares their direction of change to what is expected from the literature, in order to predict likely relevant transcriptional regulators. If the observed direction of change is mostly consistent with a particular activation state of the transcriptional regulator ('activated' or 'inhibited'), then a prediction is made about that activation state. For each TF two statistical measures are computed (overlap *P* value and activation *z*-score). The overlap *P* value labels upstream regulators based on significant overlap between data set genes and known targets regulated by a TF. The activation *z*-score is used to infer the likely activation states of upstream regulators based on comparison with a model that assigns random regulation directions. Overlap *P* value lower than 0.05 and *z*-score higher than |2| were selected as thresholds to identify a TF as relevant.

Gene length analysis. Limma was used to identify the DEGs among AL^{WT} samples compared with the other experimental conditions (DR^{WT}, AL^{Erc1} and DR^{Erc1}). Next, probe-sets in the Affymetrix array with multiple gene annotation were filtered out. BiomaRt³⁹ was used to retrieve the gene length for the remaining probe sets (32,930 probe-sets from 45,142 probe-sets in the original microarray). Differentially expressed genes were selected using an FDR of <0.05 and a linear

fold change of ± 1.5 . The Shapiro–Wilk test was applied to contrast the normality of the distribution of gene length in the different lists of DEGs. Because most of the distributions were not normal, a Mann–Whitney test for non-paired samples test was used to evaluate whether the distributions of DEGs were different between the different comparisons. Finally, a relative frequency (kernel density) plot of gene length and probability density for DEG in each comparison was drawn using the density function implemented in R. Kernel density estimates are related to histograms, but with the possibility to smooth and continuity by using a kernel function. The y axis represents the density probability for a specific range of values in the x axis.

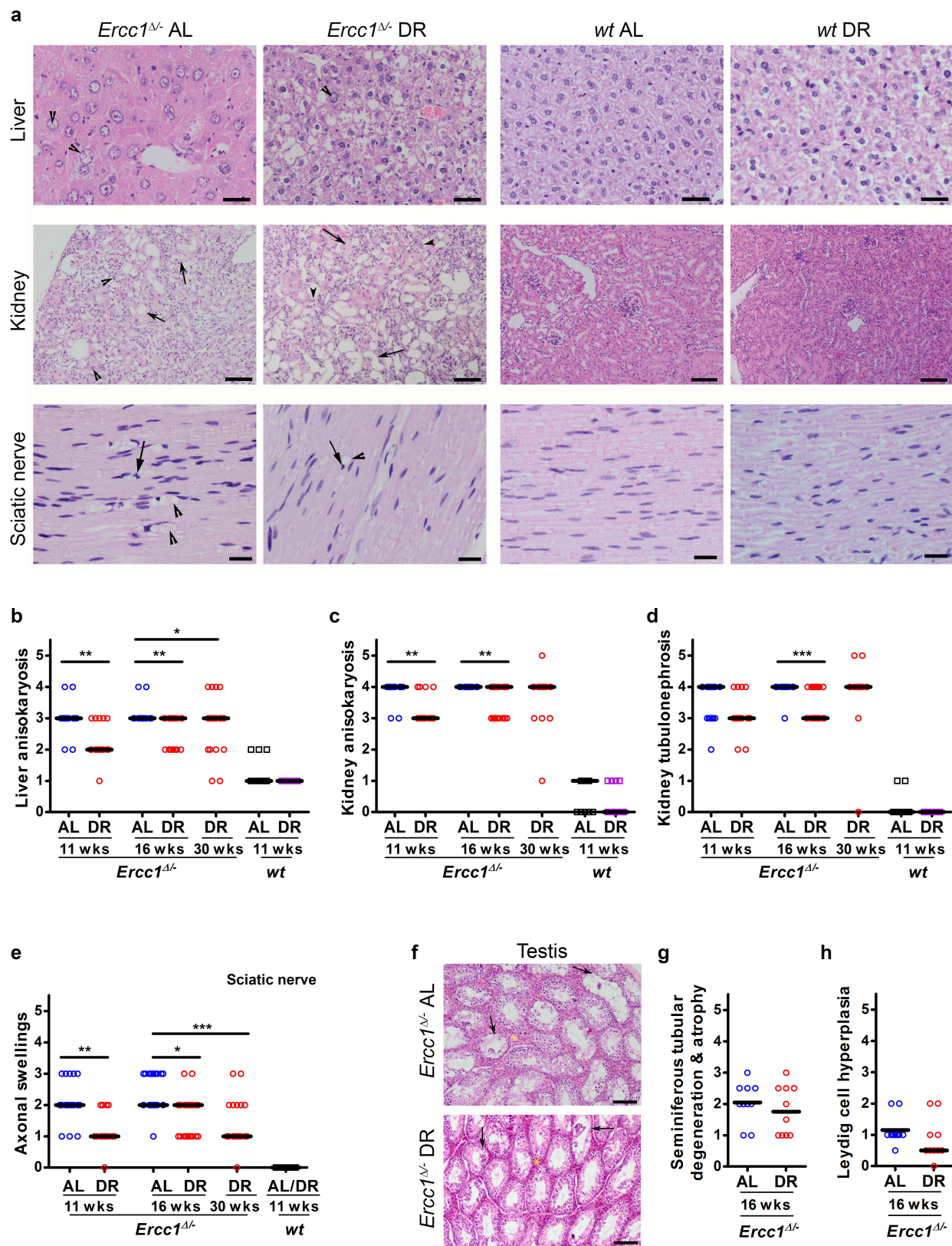
Immunoblotting. Liver extracts from *ad libitum*- and dietary restriction-treated *Ercc1*^{Δ/−} and wild type mice ($n = 6$, 11 weeks) were prepared by mechanical disruption in lysis buffer (150 mM NaCl, 1% Triton X-100, 50 mM Tris), which was supplemented with mini complete protease inhibitor (Roche Diagnostics) and phosphate inhibitors (5 mM NaF, 1 mM Na-orthovanadate). After mechanical disruption, lysates were incubated on ice for 1 h and subsequently centrifuged at 4 °C for 20 min. Lysate (25–50 μg) was loaded on a 10% SDS–PAGE gel (Life Technologies LTD) and transferred to a PVDF transfer membrane (GE-Healthcare Life Sciences). Levels of S6 (#2217S Lot5; 1:2,000), S6(Ser240/244; #2215 Lot14; 1:500), Akt (#9272 Lot25; 1:500), Akt(Ser473; #9271S Lot13; 1:250) and Akt(Thr308; #9275S Lot19; 1:500) were detected (Cell Signaling Technology), semi-quantified using the ImageJ software package (<http://rsb.info.nih.gov/ij/index.html>) and phosphorylated:total ratios relative to *ad libitum* samples were calculated. Differences between groups were assessed with a t -test. β -Actin was used as loading control (Sigma; A5441 Lot064M4789V; 1:25,000).

31. Duncan, A. W. *et al.* The ploidy conveyor of mature hepatocytes as a source of genetic variation. *Nature* **467**, 707–710 (2010).
32. McWhir, J., Selfridge, J., Harrison, D. J., Squires, S. & Melton, D. W. Mice with DNA repair gene (ERCC-1) deficiency have elevated levels of p53, liver nuclear abnormalities and die before weaning. *Nat. Genet.* **5**, 217–224 (1993).
33. Waarsing, J. H., Day, J. S. & Weinans, H. An improved segmentation method for in vivo microCT imaging. *J. Bone Miner. Res.* **19**, 1640–1650 (2004).
34. Botter, S. M. *et al.* ADAMTS5^{−/−} mice have less subchondral bone changes after induction of osteoarthritis through surgical instability: implications for a link between cartilage and subchondral bone changes. *Osteoarthritis Cartilage* **17**, 636–645 (2009).
35. Durik, M. *et al.* Nucleotide excision DNA repair is associated with age-related vascular dysfunction. *Circulation* **126**, 468–478 (2012).
36. Derks, K. W. *et al.* Deciphering the RNA landscape by RNAome sequencing. *RNA Biol.* **12**, 30–42 (2015).
37. Ringnér, M. What is principal component analysis? *Nat. Biotechnol.* **26**, 303–304 (2008).
38. Smyth, G. K., Gentleman, R., Carey, V., Dudoit, S., Irizarry, R. & Huber, W. (eds.), *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* pp. 397–420 (Springer, 2005).
39. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protocols* **4**, 1184–1191 (2009).
40. Jonker, M. J. *et al.* Life spanning murine gene expression profiles in relation to chronological and pathological aging in multiple organs. *Aging Cell* **12**, 901–909 (2013).
41. de Graaf, E. L. *et al.* Spatio-temporal analysis of molecular determinants of neuronal degeneration in the aging mouse cerebellum. *Mol. Cell. Proteomics* **12**, 1350–1362 (2013).
42. He, L. *et al.* A microRNA component of the p53 tumour suppressor network. *Nature* **447**, 1130–1134 (2007).
43. Navarro, F. & Lieberman, J. miR-34 and p53: new insights into a complex functional relationship. *PLoS One* **10**, e0132767 (2015).
44. Swindell, W. R. Genes and gene expression modules associated with caloric restriction and aging in the laboratory mouse. *BMC Genomics* **10**, 585 (2009).
45. Campisi, J. & Robert, L. Cell senescence: role in aging and age-related diseases. *Interdiscip. Top. Gerontol.* **39**, 45–61 (2014).
46. Coppé, J. P., Desprez, P. Y., Krtolica, A. & Campisi, J. The senescence-associated secretory phenotype: the dark side of tumor suppression. *Annu. Rev. Pathol.* **5**, 99–118 (2010).



Extended Data Figure 1 | Effect of dietary restriction on body weight and various healthspan parameters of *Ercc1*^{Δ/Δ} mice are primarily related to glucose metabolism and liver pathology. a–d, Body weights curves of *Ercc1*^{Δ/Δ} (a, b) and *Xpg*^{-/-} (c, d) male (a, c) and female (b, d) mice with *ad libitum* access to AIN93G diet or on 30% dietary restriction (red) shown as mean ± s.e. at weekly intervals; n = 4 animals per group, solitary housed at the EMC. Dietary restriction was initiated at 7 weeks of age at a restriction of 10% and increased weekly by 10%, until 30% was reached from 9 weeks of age onwards. e–g, Blood glucose after feeding (e), plasma fasting insulin (f), and plasma albumin levels (g), indicative of liver functioning, in *ad libitum* and dietary restriction wild-type and *Ercc1*^{Δ/Δ} mice at 16 weeks. n ≥ 3 animals per group.

h, Quantification of 16N nuclei in hepatocytes³² of 11-week-old male wild-type and *Ercc1*^{Δ/Δ} mice under *ad libitum* or dietary restriction regimens by FACS analyses; n = 5 animals per group. i, Total numbers of splenic CD4⁺ T cell from spleen of 16-week old *Ercc1*^{Δ/Δ} mice under dietary restriction or *ad libitum* and aged-matched wild-type controls. n ≥ 3 animals per group. j, IgA blood levels in male *Ercc1*^{Δ/Δ} mice at different ages under dietary restriction or *ad libitum* regimes. n = 5 animals per group. k, Average grip strength of the forelimbs and all limbs of 16-week old *Ercc1*^{Δ/Δ} and wild-type mice is similar under *ad libitum* and dietary restriction conditions; n = 4 animals per group. Mean ± s.e. *P < 0.05, **P < 0.01, ***P < 0.001.

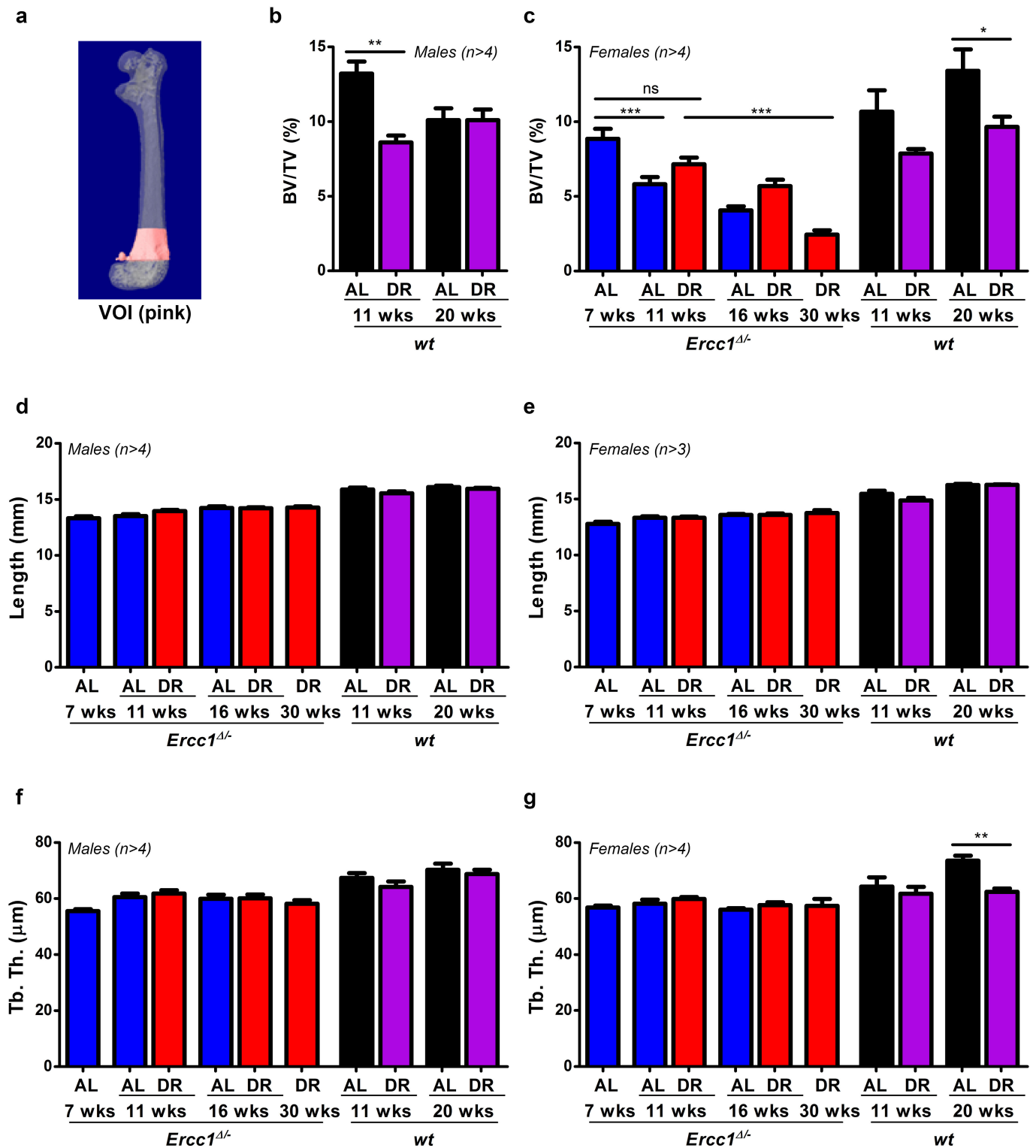


Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | Dietary restriction improves ageing-related histopathological phenotypes in different tissues of *Ercc1*^{Δ/−} mice.

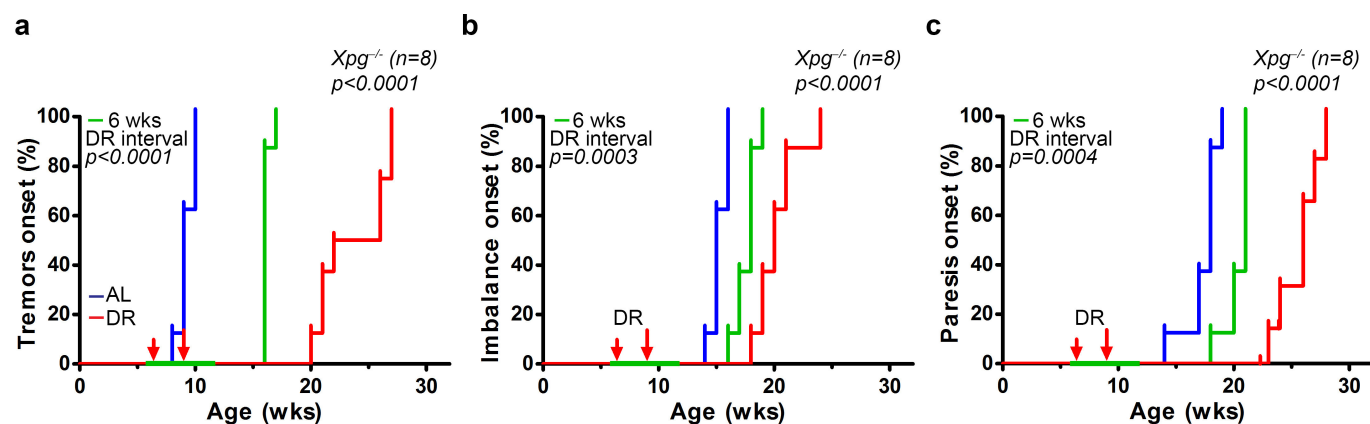
a, Representative pictures of haematoxylin-eosin-stained slides from liver, kidney, and sciatic nerve. From left to right: AL^{Ercc1}, DR^{Ercc1}, AL^{WT} and DR^{WT}. Lesions were semiquantitatively assessed, with scores ranging from absent (0) to massive (5). The liver of a female AL^{Ercc1} mouse shows moderate anisokaryosis (score = 3) and intranuclear inclusions (score = 3, arrowheads). The liver of a female DR^{Ercc1} mouse shows moderate hydropic degeneration with mild anisokaryosis (score = 1) and a few hepatocellular intranuclear inclusions (score = 1, arrowhead). Histologically normal liver tissue from AL^{WT} and DR^{WT} mice. The kidney of a female AL^{Ercc1} and DR^{Ercc1} mouse with severe tubular attenuation and degeneration (score = 5, arrows) with marked anisokaryosis (score = 4, arrowheads) next to histologically normal kidneys from female AL^{WT} and DR^{WT} mice. The sciatic nerve of a female AL^{Ercc1} mouse with severe axonal swellings (score = 3, arrowheads). These axonal swellings probably represent vacuoles containing myelin debris and/or fragmented axons. The Schwann cell nuclei around vacuolated areas are pyknotic (arrow). The sciatic nerve of a female DR^{Ercc1} mouse displays mild vacuole-like structures (score = 1, arrowhead) with pyknosis of Schwann cell nuclei (arrow), while the histologically normal sciatic nerves of female AL^{WT} and DR^{WT} mice display no axonal swellings. Scale bar in liver, 50 μm; in kidney, 100 μm; in sciatic nerve, 20 μm. **b**, Pathology assessment of anisokaryosis in livers from *Ercc1*^{Δ/−} mice at different ages under *ad libitum* (blue) or dietary restriction (red) regimen and young *ad libitum* (black) and dietary restriction (purple) wild-type controls. Scores range from absent (0) to

massive (5); $n \geq 10$ animals per group; bars indicate group medians. **c, d**, Pathology assessment of anisokaryosis (**c**) and tubulonephrosis (**d**) in the kidneys of *Ercc1*^{Δ/−} and wild-type mice at different ages under *ad libitum* and dietary restriction regimes. Scores range from absent (0) to massive (5); $n \geq 10$ animals per group. **e**, Pathology assessment of axonal swellings in sciatic nerves of *Ercc1*^{Δ/−} mice at different ages under *ad libitum* or dietary restriction regimes. Scores range from absent (0) through massive (5); $n \geq 10$ animals per group. **f**, Representative pictures to the testicular lesions observed in *Ercc1*^{Δ/−} males. The AL^{Ercc1} testes (upper panel) exhibited moderate testicular degeneration and atrophy (arrows). Also, the Leydig cells (yellow asterisk) appeared more prominent, probably owing to the tubular loss and attenuation, or possibly owing to true Leydig cell hyperplasia (a common ageing lesion in rodent testes). These phenotypes were slightly rescued in the testes of DR^{Ercc1} mice (lower panel). **g, h**, Pathology assessment of seminiferous tubular degeneration and atrophy (**g**) and Leydig cell hyperplasia (**h**) in the testes of *Ercc1*^{Δ/−} mice at 16 weeks of age under *ad libitum* (blue) or dietary restriction (red) regimen. Scores were given as absent (0), subtle (1), mild (2), moderate (3), severe (4), and massive (5) for each criteria, with a 0.5 interval; $n = 10$ animals per group; bars indicate group medians. Note that testicular development is mostly completed at the start of dietary restriction. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. The values for the wild-type mice do not change significantly in the timeframes used here (see ref. 40). Pathological scores, including those of other liver and kidney ageing-related histopathological phenotypes, are given in Supplementary Table 1.



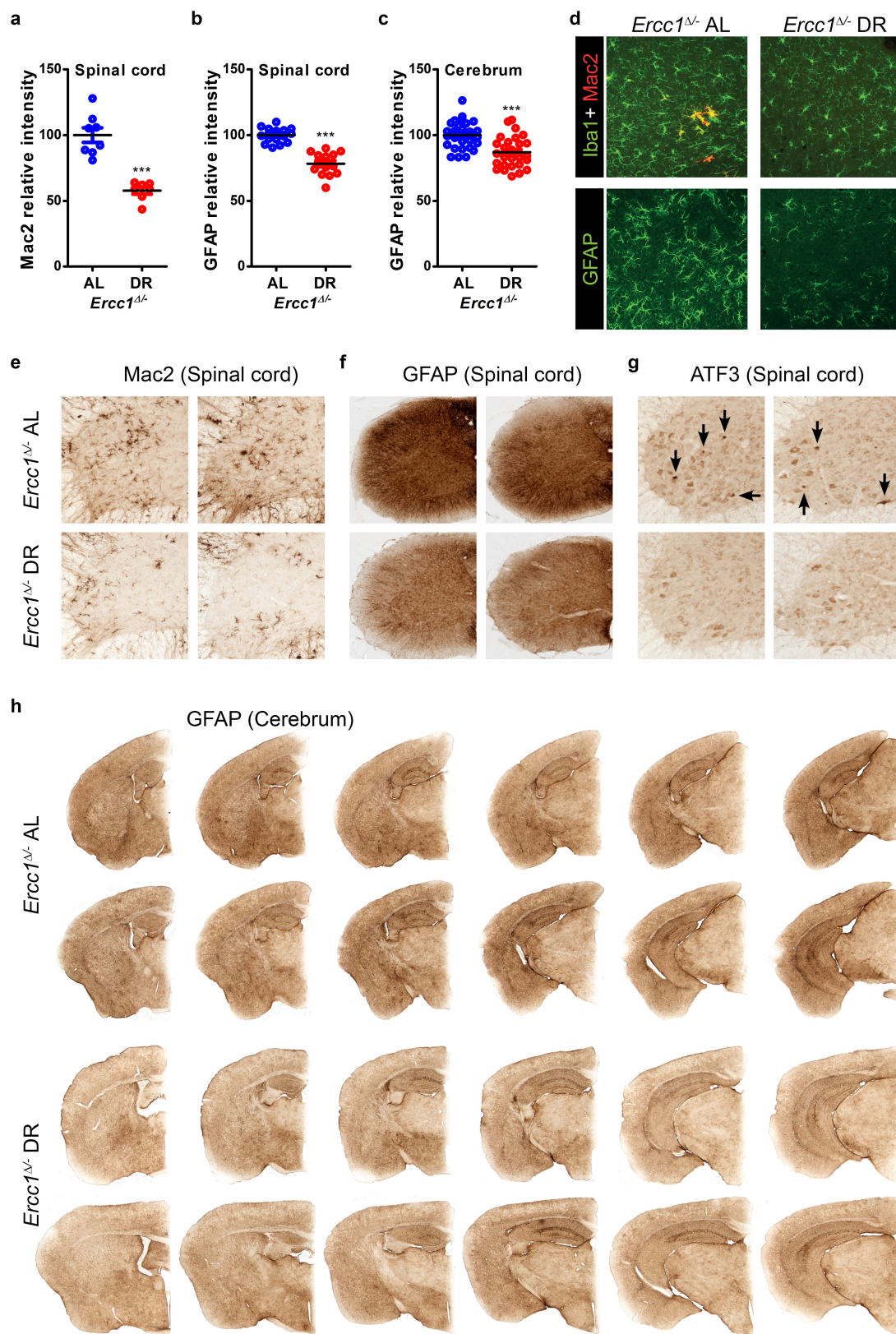
Extended Data Figure 3 | Dietary restriction preserves skeletal structure in *Ercc1*^{Δ/−} mice. **a**, Illustration depicting the femoral volume of interest (VOI) for microCT analyses. **b**, **c**, Trabecular bone volume fraction (BV/TV) representing the amount of trabecular bone in the femur VOI of wild-type male mice (**b**) as well as *Ercc1*^{Δ/−} and wild-type female mice (**c**) expressed as percentage measured using micro-CT. **d**, **e**, Femur length of *Ercc1*^{Δ/−} and wild-type male (**d**) and female (**e**) mice.

f, **g**, Trabecular thickness in the femur VOI of *Ercc1*^{Δ/−} and wild-type male (**f**) and female (**g**) mice. *Ad libitum*- and dietary restriction-treated animals were measured at different ages with $n \geq 3$ animals per group. Values of *Ercc1*^{Δ/−} mice are depicted in blue (*ad libitum*) and red (dietary restriction). Young wild-type controls are depicted in black (*ad libitum*) and purple (dietary restriction). Error bars denote mean \pm s.e. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.



Extended Data Figure 4 | Dietary restriction preserves neurofunctional behaviour of $Xpg^{-/-}$ mice. a–c, Onset of neurological abnormalities as tremors (a), imbalance (b), and paresis of the hind limbs (c) with age in $Xpg^{-/-}$ mice under *ad libitum* and dietary restriction regimens. $n = 8$ animals per group. The onset of continuous dietary restriction is indicated by the red arrows. Average age at the onset of tremors is delayed from 9 to

24 weeks, imbalance from 15 to 20 weeks, and paresis from 18 to 26 weeks. Temporary dietary restriction was given between 6 and 12 weeks of age and is indicated in green. This short period of dietary restriction yielded a median delay in onset of tremors of 7 weeks while the median age of onset of both imbalance and paresis was delayed by 3 weeks. P values were calculated against $Xpg^{-/-}$ -*ad libitum* mice using the log-rank test.

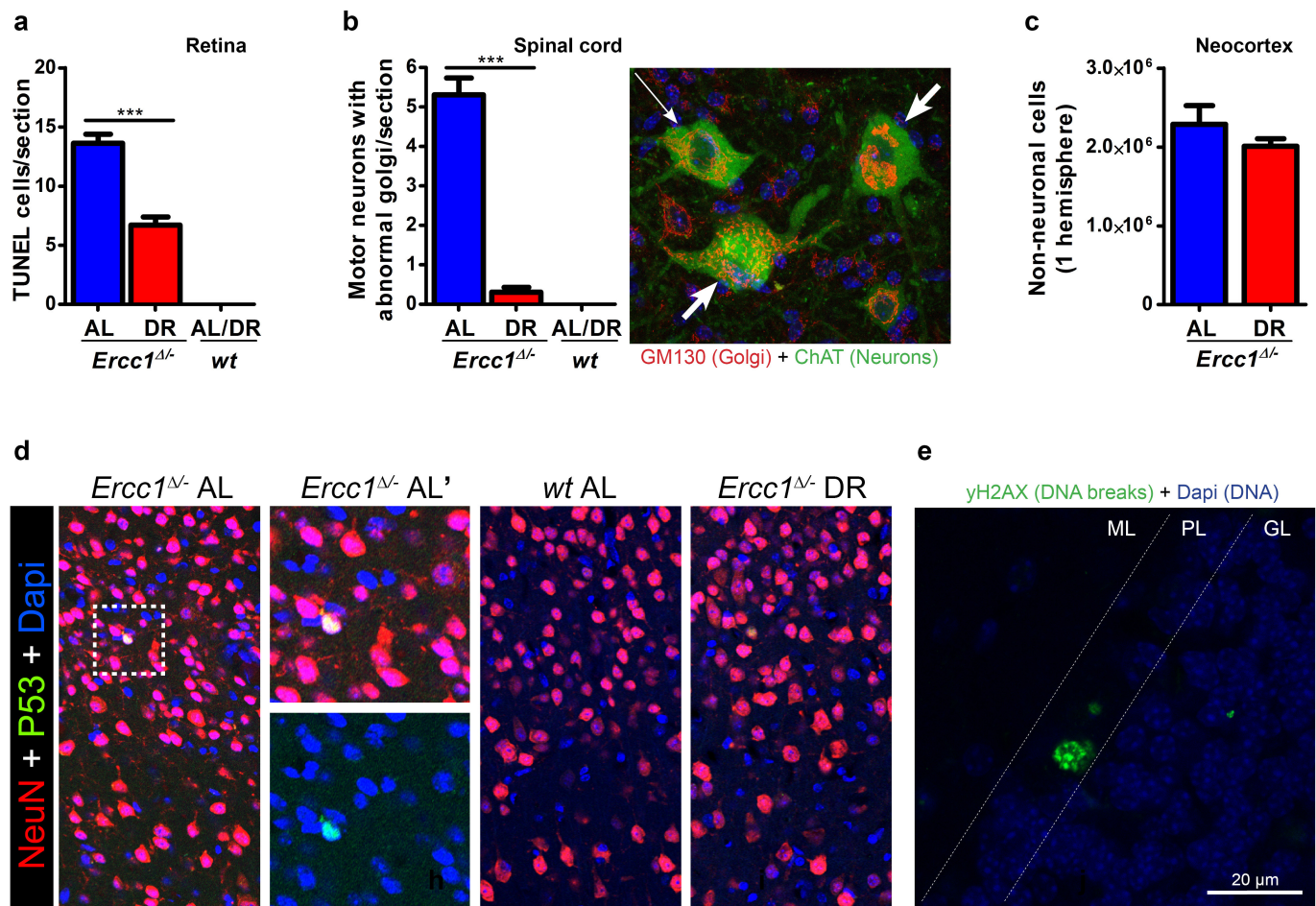


Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | Dietary restriction improves microgliosis and astrocytosis in brain and spinal cord of *Ercc1*^{Δ/−} mice. a–c,

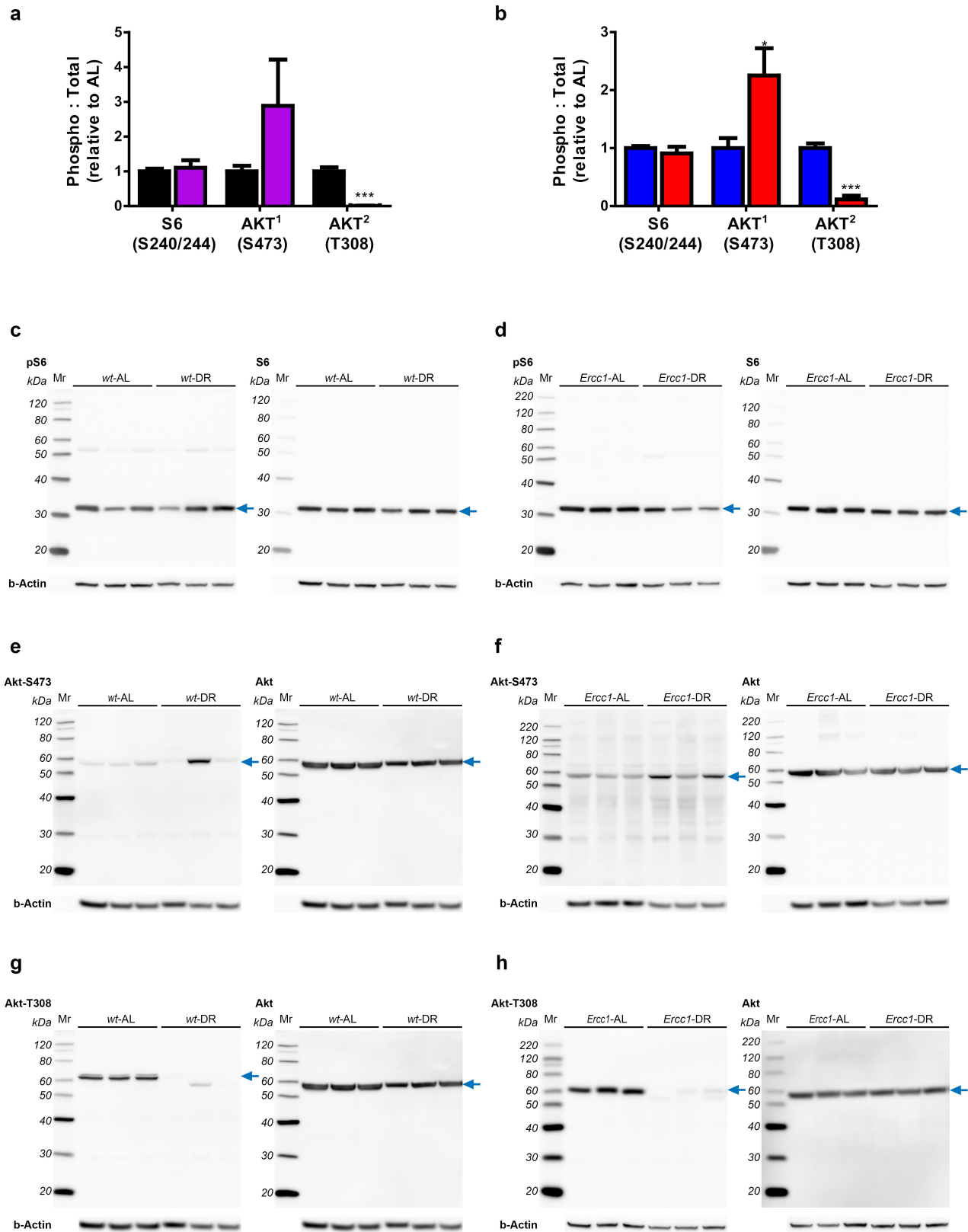
Quantification of the relative intensity of consecutive transverse brain and spinal cord sections immunoperoxidase-stained for Mac2 in spinal cord (a) and GFAP in spinal cord (b) and cerebrum (c). $n > 3$ animals per group; bars indicate group medians. d, Iba1, Mac2, and GFAP immunofluorescent confocal images showing that reduced astrocytosis (GFAP) in cortex is paralleled by reduced staining for microglia (Iba1). Mac2-immunoreactivity, which outlines a subset of phagocytosing microglia cells, is also reduced in the neocortex of 16-week-old DR^{Ercc1} mice ($n = 4$) when compared to *ad-libitum* mice ($n = 3$). e–g, Representative pictures of spinal cord sections of 16-week-old AL^{Ercc1} and DR^{Ercc1} mice immunoperoxidase-stained for Mac2 (e) and GFAP

(f) reflecting reduced microgliosis and astrocytosis, respectively, in the nervous system of diet-restricted mice. Immunoperoxidase-stained spinal cord sections for ATF3 (g) showed that activation of the stress-inducible transcription factor ATF3 (which is induced following genotoxic stress via p53-dependent and -independent pathways) is less pronounced in the nervous system of diet-restricted mice. Sections from two different animals are presented next to each other. Black arrows indicate cells with high nuclear ATF3 staining. h, Representative pictures of consecutive transverse brain sections of 16-week-old AL^{Ercc1} and DR^{Ercc1} mice immunoperoxidase-stained for GFAP, showing reduced GFAP staining in the nervous system of diet-restricted mice. Six 40 μm slices are shown per animal, with 360 μm cerebrum thickness between each slice. Mean ± s.e. *** $P < 0.001$.



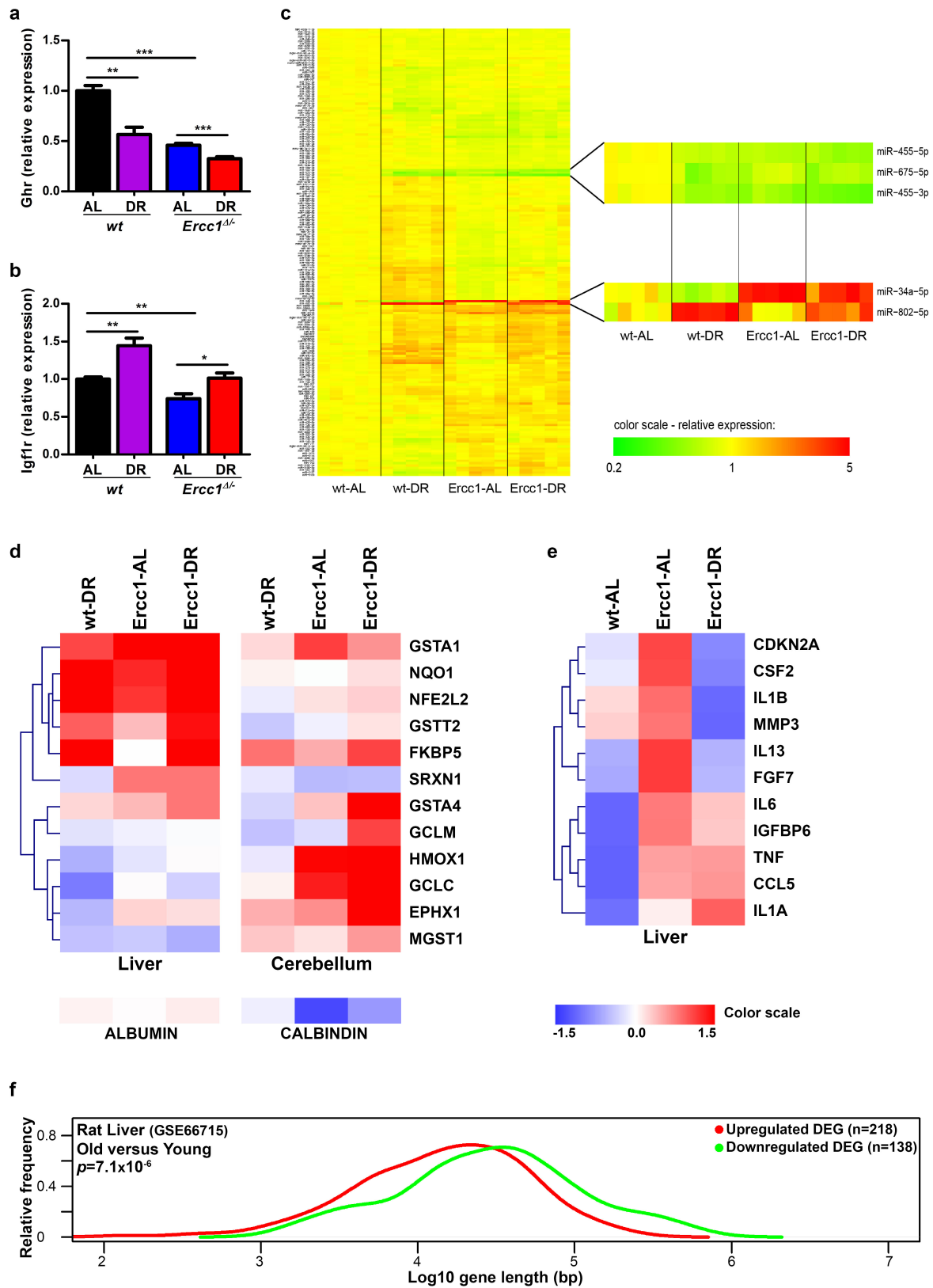
Extended Data Figure 6 | Dietary restriction dramatically preserves neurofunctioning of *Ercc1*^{Δ/Δ} mice. **a**, Quantification of TUNEL-positive cells in the outer nuclear layer of retinal sections of 16-week-old *ad libitum* (blue) or diet-restricted (red) *Ercc1*^{Δ/Δ} mice; *n* = 4 animals per group. **b**, Analysis of the total number of motor neurons with abnormal Golgi apparatus (indicative of impaired cells, see thick arrows in representative image; neuron with normal Golgi is indicated by a thin arrow) in C6 cervical spinal cord sections from 16-week-old diet-restricted and *ad libitum* *Ercc1*^{Δ/Δ} mice. *n* = 4 animals per group. TUNEL-positive cells (**a**) and neurons with abnormal Golgi morphology (**b**) were absent in both *ad libitum*¹⁷ and diet-restricted young wild-type mice. **c**, Quantitative stereological analysis of the total number of non-neuronal cells (DAPI⁺/NeuN⁻; *P* = 0.2744) in the neocortex of transverse brain sections of 16-week-old *ad libitum* and diet-restricted *Ercc1*^{Δ/Δ} mice. *n* ≥ 3 animals per group. Mean ± s.e. ****P* < 0.001. **d**, Representative images of

neocortex stained for NeuN (neurons), p53 and DAPI (for staining DNA) used for quantitative stereological analysis of the total number of neurons (NeuN⁺) and non-neuronal cells (DAPI⁺/NeuN⁻) in 16-week old *ad libitum*- (*n* = 3) and dietary restriction- (*n* = 4) treated *Ercc1*^{Δ/Δ} mice. Quantification of the number of p53-positive neurons is shown in Fig. 3e. The analysis was performed using the optical dissector probe from StereoInvestigator on a Zeiss LSM700 laser-scanning microscope. **e**, Representative image of cerebellum stained for γH2AX (green, double-stranded DNA breaks) and DAPI (blue, for staining DNA) in 16-week-old *ad libitum* (*n* = 3) and diet-restricted (*n* = 4) *Ercc1*^{Δ/Δ} mice. The Purkinje (PkJ) neurons are present in a single layer (PL, the purkinje layer) in between the molecular layer (ML) and granular layer (GL)⁴¹. Quantification of the number of γH2AX-positive PkJ-neurons is shown in Fig. 3h. The analysis was performed using a Zeiss LSM700 laser scanning microscope.



Extended Data Figure 7 | Effect of dietary restriction on mTorC1, mTorC2 and Ins/PDK1 signalling using immunoblot analysis of wild-type and *Ercc1*^{Δ/−} liver extracts. **a, b**, Quantified relative S6 and Akt phosphorylation by dietary restriction in wild-type (**a**) and *Ercc1*^{Δ/−} (**b**) liver extracts. Six animals per group were used at 11 weeks of age. **c–h**, Representative images used to quantify the ratio of S6 and Akt phosphorylation versus total S6 and Akt respectively. Phosphorylation of

Akt at position S473 seems to be increased by dietary restriction in liver homogenates of 11-week-old wild-type (**e**) and *Ercc1*^{Δ/−} (**f**) mice, but is suppressed at position T308 (**g, h**). Phosphorylation of S6 at S240 and S244 is unaffected by dietary restriction (**c, d**). For immunoblots, data for three animals per group are shown. For graphs and statistics, six animals per group were used. The blue arrow indicates signals used for quantification. Below each blot, β -actin is presented as a loading control.



Extended Data Figure 8 | See next page for caption.

Extended Data Figure 8 | Molecular analysis of expression changes by diet, DNA damage, or ageing. **a, b**, *Ghr* and *Igf1r* gene expression changes measured by quantitative real-time PCR (qRT-PCR) in liver samples of 11-week old wild-type and *Ercc1*^{Δ/−} mice with restricted diets ($n = 5$). Gene-specific real-time PCR primers are described in Methods. **c**, MicroRNA expression profile comparison of wild-type and *Ercc1*^{Δ/−} mouse liver tissue under *ad libitum* and diet-restricted conditions. Shown are 188 significantly regulated miRNAs (FRD $\leq 5\%$) between groups. Five of the most significantly changed microRNAs are zoomed in. miR-34a, a downstream target of p53 that is involved in cell cycle regulation and apoptosis, is induced by DNA damage^{42,43}. It showed differential expression between liver homogenates of 11-week-old wild-type and *Ercc1*^{Δ/−} mice. It was downregulated by dietary restriction in the liver of wild-type mice (1.62 fold, $P = 0.02$), but strongly upregulated in the liver of AL^{Ercc1} mice compared to AL^{WT} mice (4.7-fold, $P = 0.0001$) and seems suppressed in DR^{Ercc1} expression profiles. These changes were confirmed by qPCR (data not shown). **d**, Heat map of key antioxidant defence genes in liver and brain of wild-type and *Ercc1*^{Δ/−} mice. Fold changes were calculated for DR^{WT}, AL^{Ercc1}, and DR^{Ercc1} mice against AL^{WT} mice, using microarray expression profiles of liver tissue at 11 weeks of age ($n = 5$) or qRT-PCR for cerebellum tissue at 16 weeks of age ($n = 4$). Dietary restriction induced an antioxidant response in liver, which is less pronounced in brain specimens, consistent with earlier findings⁴⁴. This is likely to be due to the high endogenous antioxidant defence levels in the nervous system. The difference in antioxidant response between liver and brain by genotype conforms to previous results⁶. Interestingly, the Purkinje neuron marker calbindin is clearly reduced in cerebella of AL^{Ercc1} mice but is less reduced in DR^{Ercc1} mice, confirming the strong reduction in DNA-damage-induced Purkinje cell loss induced by dietary restriction. Blue, decreased expression; red, increased expression. Hierarchical clustering on liver and cerebellum genes was performed using a Pearson correlation. **e**, Dietary restriction reduces the p16-RB branch of senescence and the senescence-associated secretory phenotype (SASP)

as assessed by next-generation sequencing expression analysis of the liver RNA of *Ercc1*^{Δ/−} mice. To assess the p16-RB branch of the senescence phenotype⁴⁵, we followed a next-generation sequencing approach as previously described³⁶ using 16-week-old liver tissue from AL^{WT}, DR^{WT}, AL^{Ercc1} and DR^{Ercc1} mice ($n = 1$). By sequencing >150M sequence reads per sample, we detected the *p16-ink4a* (*Cdkn2a*) transcript at sufficient levels. *p16-ink4a* (*Cdkn2a*) is considered a key marker for cellular senescence, but is difficult to quantitatively analyse using other methods owing to the high ratio of normal cells to senescent cells. Data sets were normalized by calculating reads per kilobase million (RPKM). Subsequently, z-scores were calculated and plotted in a heat map. Red, increased expression; blue, decreased expression. In AL^{Ercc1} liver RNA, *p16-ink4a* (*Cdkn2a*) is upregulated compared to levels in AL^{WT} animals, but downregulated after dietary restriction. This indicates that *Ercc1*^{Δ/−} mice have increased cellular senescence that is reduced upon dietary restriction. Second, we monitored the transcriptionally induced SASP as described previously⁴⁶. Many, if not all, SASP factors are not exclusively specific for cellular senescence. To reduce the probability that observed SASP factor expression changes are contributed to by other cells, we selected only those SASP factors that have an absolute expression (RPKM) in the same range as *p16-ink4a* in across these data sets, since these are most likely to be the result of cellular senescence. The figure shows that most SASP factors such as *IL-6*, the most prominent SASP cytokine are downregulated after dietary restriction. This supports the idea that cellular senescence and associated SASP are increased in AL^{Ercc1} liver and are reduced by dietary restriction. Hierarchical clustering was performed using a Pearson correlation. **f**, Suppression of long genes in normal ageing of rat liver. A relative-frequency plot of the gene length of DEGs in liver tissue from 24-month old rat versus that seen in 6-month old rat. Upregulated genes, red; downregulated genes, green. The DEGs from rat liver were selected using a fold-change cut-off of 1.5 and an FDR <0.05. The data set is publicly available in the NCBI Gene Expression Omnibus under accession number GSE66715.

Extended Data Table 1 | Discordant DEG in dietary restriction response for wild-type and *Ercc1*^{Δ/-} mice

ProbeID	Gene Symbol	logFC wt-DR vs wt-AL	P.Val Wt-DR vs wt-AL	logFC Ercc1-DR vs Ercc1-AL	P.Val Ercc1-DR vs Ercc1-AL
1438583_PM_at	Ern1	0.77	0.003	-0.74	0.010
1438997_PM_at	Ern1	0.69	0.005	-0.70	0.010
1429295_PM_s_at	Trip13	-0.72	0.005	0.77	0.007
1441098_PM_at	Pnldc1	-0.66	0.030	1.10	0.002

DEG in common in dietary restriction response in liver of wild-type and *Ercc1*^{Δ/-} mice but with discordant direction of change.

Extended Data Table 2 | Average gene length of up- and downregulated genes of wild-type and *Ercc1*^{Δ/-} liver expression profiles under *ad libitum* and diet-restricted conditions

Comparison	Length (bp) means	wilcox.test
Wt-DR Up vs. Ercc1-AL Up	66,364 vs. 40,021	< 2.2e-16
Wt-DR Down vs. Ercc1-AL Down	62,352 vs. 212,957	< 2.2e-16
Wt-DR Up vs. Ercc1-DR Up	66,364 vs. 41,284	< 2.2E-16
Wt-DR Down vs. Ercc1-DR Down	62,352 vs. 141,391	3.24E-14
Ercc1-AL Up vs. Ercc1-DR Up	40,021 vs. 41,284	0.001
Ercc1-AL Down vs. Ercc1-DR Down	212,957 vs. 141,391	0.0002

Comparisons of the mean gene size distributions of DEGs (up- and downregulated) among DR^{WT} versus AL^{WT}, AL^{Ercc1} versus AL^{WT} and DR^{Ercc1} versus AL^{WT} in mouse liver tissue. wt-DR Up, shows the number of DEGs that are upregulated in DR^{WT} compared with AL^{WT} ($n = 1,106$); wt-DR Down shows the number of DEGs that are downregulated in DR^{WT} compared with AL^{WT} ($n = 1,046$); Ercc1-AL Up shows the DEGs that are upregulated in AL^{Ercc1} compared with AL^{WT} ($n = 595$); Ercc1-AL Down shows the DEGs that are downregulated in AL^{Ercc1} compared with AL^{WT} ($n = 363$); Ercc1-DR UP shows the number of DEGs that are upregulated in DR^{Ercc1} compared with AL^{WT} ($n = 1,384$); Ercc1-DR Down shows the number of DEGs that are downregulated in DR^{Ercc1} compared with AL^{WT} ($n = 768$). Because the distributions of the gene lengths do not have a normal distribution the means should be taken as a reference value.

CORRIGENDUM

doi:10.1038/nature18269

Corrigendum: Convergence of terrestrial plant production across global climate gradients

Sean T. Michaletz, Dongliang Cheng, Andrew J. Kerkhoff & Brian J. Enquist

Nature **512**, 39–43 (2014); doi:10.1038/nature13470

It has come to our attention that in this Article, while translating the methods of Luo¹ (originally written in Chinese), we did not appreciate that plant age (a) and stand biomass (M_{tot}) had been used to calculate net primary production (NPP). Thus, while the Luo data are appropriate for our analyses that used climate and environmental variables as predictors of NPP, they are not appropriate for those that use plant age and/or stand biomass as independent predictors as in our theoretical model.

Consequently, we have removed the Luo data (our data index numbers 98–1206) from all analyses that involve age and biomass as predictors. This error has been corrected in the Supplementary Information to this Corrigendum, which contains revised versions of Table 1, Figs 1d, 3, 4, Extended Data Tables 2, 3, and Extended Data Figs 1, 3, 4. These revisions are based on the subset of our original source data file that excludes ‘Source’ rows containing ‘Luo (1996); Ni et al. (2001)’ (see Supplementary Information to this Corrigendum for the corrected source data file). We also include the first English translation and summary of the methodology from Luo¹, as previous studies have made the same error by correlating NPP with age or biomass using the Luo data^{2–5}.

Overall, this correction does not change our original interpretation of the results or the conclusions drawn in the Article. Furthermore, there is little effect on the parameter fits reported in our original Article. A few differences merit discussion here. First, our re-analyses strengthen our conclusions that growing season length (l_{gs}) is an important indirect driver of variation in NPP, as well as our rationale for calculation of growing season net primary production NPP/l_{gs} . (The corrected

Table 1 in the Supplementary Information to this Corrigendum shows that l_{gs} explains an even larger fraction of NPP than it did in our original analysis.) Second, our fitted estimates for activation energy (E) for NPP and NPP/l_{gs} (see corrected Table 1 in Supplementary Information to this Corrigendum) now have 95% confidence intervals that include the value of 0.32 eV proposed for photosynthesis⁶. This result should be interpreted with caution, however, given that the confidence intervals are wide and temperature still explains a relatively small amount of variation in NPP and NPP/l_{gs} . Third, the fitted mass-scaling exponents α for total NPP/l_{gs} and the aboveground woody component $\text{NPP}_{\text{AGW}}/l_{\text{gs}}$, while similar, have switched places in their correspondence to theoretical predictions of 0.6, with $\alpha = 0.47$ for NPP/l_{gs} (95% confidence interval = 0.36–0.58; see corrected Table 1 in the Supplementary Information to this Corrigendum) and $\alpha = 0.552$ for $\text{NPP}_{\text{AGW}}/l_{\text{gs}}$ (95% confidence interval = 0.374–0.729; see corrected Extended Data Table 3 in the Supplementary Information to this Corrigendum). A closer correspondence for $\text{NPP}_{\text{AGW}}/l_{\text{gs}}$ may be expected due to possible bias in sampling below ground biomass for estimates of NPP. We apologize for any confusion that this oversight may have caused to readers.

We are grateful to B. Medlyn, J. Yang, and their journal club at Western Sydney University for bringing this issue to our attention. We also thank T. Luo, J. Ni and Z. Hu for helping us translate the original Chinese publication.

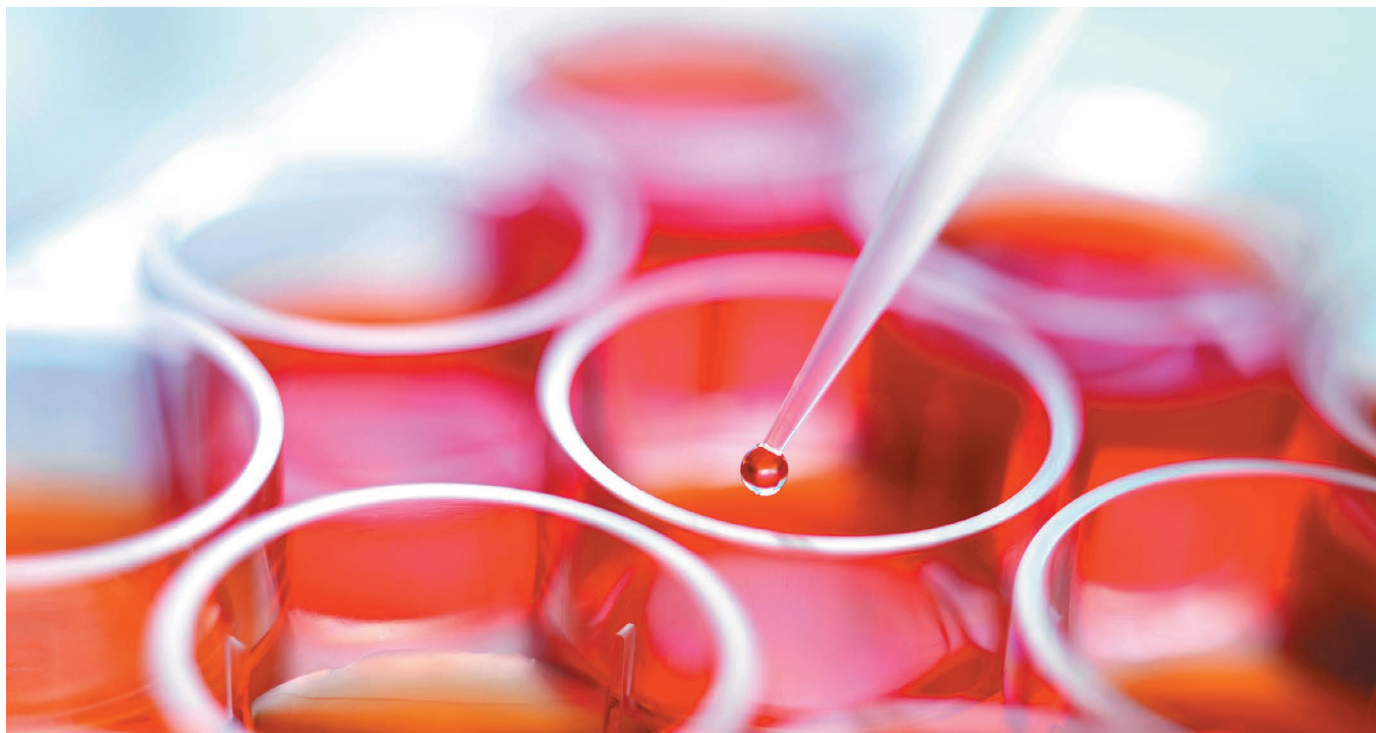
Supplementary Information is available in the online version of the Corrigendum.

1. Luo, T. X. *Patterns of Biological Production and its Mathematical Models for Main Forest Types of China*. PhD thesis, Chinese Academy of Sciences (1996).
2. Luo, T., Li, W. & Zhu, H. Estimated biomass and productivity of natural vegetation on the Tibetan Plateau. *Ecol. Appl.* **12**, 980–997 (2002).
3. Ni, J., Zhang, X. S. & Scurlock, J. M. O. Synthesis and analysis of biomass and net primary productivity in Chinese forests. *Ann. For. Sci.* **58**, 351–384 (2001).
4. Li, H.-T., Han, X.-G. & Wu, J.-G. Variant scaling relationship for mass-density across tree-dominated communities. *J. Integr. Plant Biol.* **48**, 268–277 (2006).
5. Hui, D., Wang, J., Le, X., Shen, W. & Ren, H. Influences of biotic and abiotic factors on the relationship between tree productivity and biomass in China. *For. Ecol. Manage.* **264**, 72–80 (2012).
6. Allen, A. P., Gillooly, J. F. & Brown, J. H. Linking the global carbon cycle to individual metabolism. *Funct. Ecol.* **19**, 202–213 (2005).

REPRODUCIBILITY: RESPECT YOUR CELLS!

Numerous variables can torpedo attempts to replicate cell experiments, from the batch of serum to the shape of growth plates. But there are ways to ensure reliability.

TEK IMAGE/SPL



Subtle aspects of cell culture can wreck results. Researchers should check cell identity and behaviour, and carefully characterize reagents.

BY MONYA BAKER

When Alastair Khodabukus tried to engineer muscle fibres in his new laboratory, he saw something strange: the tissue was convulsing. He had been growing fibres from the same mouse-derived clone for years, but these were different. They burned more glucose, contained lower amounts of a protein that promotes faster relaxation and fatigued less readily than those he had grown before his lab moved from Dundee, UK, to the University of California, Davis. The difference, he thinks, was due to how cows are raised in the United States¹.

Most academic labs culture cells by using fetal bovine serum (FBS), a liquid extracted from clotted cow blood and collected from abattoirs when pregnant cows are slaughtered. What ends up in the serum depends on factors

such as diet, geographical location, time of year, whether the animals receive hormones or antibiotics and the gestational age of fetal calves. Substantial amounts of FBS are added as a supplement to the culture media in which cells grow; 5–15% of the volume of growth media is typical. FBS composition can affect how thick an engineered tissue becomes, cause spontaneous artefacts that mimic cell activity and even influence how surface receptors respond to a given compound. “FBS is like a big dark cloud over our heads, not knowing what’s real and what’s not,” says Khodabukus, now a postdoctoral researcher at Duke University in Durham, North Carolina.

And serum is just one of many factors that researchers have to consider when studying cells. At a US National Institutes of Health (NIH) workshop on cell culture and reproducibility last year, Richard Neve, a cancer

biologist at the biopharmaceutical company Gilead Sciences in Foster City, California, worried that researchers could become overwhelmed. “A lot of labs see the magnitude of the problem and the complexity of the problem, and enter the primordial part of their brain and shut down.” With the right mindset, however, and some obsessive checking and planning, researchers can gain confidence in performing their experiments.

The most basic step is to ensure cells’ genetic identity. Journals and funders now ask researchers to disclose whether they have checked to make sure that, say, cell lines representing corneal or skin tissue are not actually a fast-growing line derived from human cervical cancer. But cells’ behaviour can also change with density, proliferation rates, growth media, the presence of contaminants and the time kept in culture². ►

► Serum is arguably the most common supplement in cell-culture media, and also the least consistent. Human serum harbours thousands of distinct proteins originating from a wide range of cells and tissues, as well as thousands of small-molecule metabolites, all in varying concentrations. FBS probably has similar complexity, with plentiful factors to support a fast-growing fetus, too.

FBS is not only variable, it also differs from the fluid that cells are exposed to in their natural environment. Most cells are in contact not with blood directly but with the interstitial fluid that bathes organs, says Adam Elhofy, chief science officer at Essential Pharmaceuticals in Ewing, New Jersey, a company developing a serum replacement for multiple cell types. Hormones, growth factors and other signalling molecules are abundant in serum, but tightly regulated in organs, he says (see 'Bovine serum's wide range').

GOING SERUM-FREE

To overcome such concerns, reagent firms have developed serum-free growth media. Scientists pursuing 'bioprocessing' applications — such as the manufacture of therapeutic proteins and vaccines, a process in which animal products are frowned on — have embraced the serum-free alternative. Stem-cell researchers, who know these cells are sensitive to even small changes in growth conditions, are also enthusiasts.

Many more researchers are now beginning to pay attention to how they treat their cells, driven by concerns about consistency and a push into translational medicine. These priorities are encouraging more scientists to avoid serum, says Ken Yoon, who is head of strategic marketing in the research division of MilliporeSigma, a life-science reagents company in Billerica, Massachusetts. Chemically defined, serum-free media is one of the fastest growing segments in the cell-culture space, he adds.

But serum-free media are not always possible, or pragmatic. "Everyone agrees it would be a great thing if we can move away from FBS and to something more defined," says Jon Lorsch, head of the US National Institute of General Medical Sciences in Bethesda, Maryland. "The question is how feasible it is, and we don't know the answer to that question."

Most serum-free formulations apply only to a specific cell type or closely related group of cell lines. Vendors sell one serum-free medium for, say, Chinese hamster ovary cells, an epithelial cell line that is often used to produce therapeutic proteins, and others to expand particular types of blood cells. Formulations don't work for all cell types: many 'primary cells' — those taken directly from living tissue — require serum to grow after they are removed from the cues that the body provides, says Jennifer Welser-Alves, associate director of research and development at ScienCell Research Laboratories in Carlsbad, California.

"Anything you can do to boost the cells and keep them growing is necessary," she says. Some formulations require adding just 2% FBS to primary growth media, a low volume of serum that helps cut down on variability.

Even if the option is available, many researchers are unwilling to take the time, or the risk, to wean their cells off serum, says Paul Price, a culture-media consultant in Mount Pleasant, South Carolina, who has designed serum-free formulations. "Every year since 1980, people have been saying that serum is dead," he says. "Serum is still very popular because people like the idea that they can grow cells and not have fabulous technique." Culture is tough on cells: researchers pipette them from dish to dish, freeze and thaw them, add digestive enzymes to detach them from substrates and more. Serum is a balm for such abuses, says Price.

STUCK WITH SERUM

No commercial formulation is available for skeletal muscle fibres, says Khodabukus. He has spent two years tinkering with recipes that combine dozens of growth factors and other signalling molecules. When the fibres' performance changes with each lot of serum, it disrupts his own projects and muddies collaborations, he says. "I'm going to spend the rest of my life working with this system, and as a scientist I want control. If we can get this to work and be consistent, we can get this to work in every lab around the world."

Keith Baar, Khodabukus's former post-doc adviser at the University of California, Davis, relies on a more common solution: he keeps a freezer in his lab that's dedicated to

storing serum. When serum starts to run low, he orders and tests at least four batches, and watches the cells' performance to find the closest match to that in his current experiments. Then he buys 100 bottles from the same lot of serum. That can drain US\$25,000 from his lab

"One of the hardest things to assess is what constitutes a healthy cell."

budget, but it means that his lab members can continue their experiments without stopping every few months to test more lots of serum.

Researchers who don't test their serum could run into trouble, says Matthew Sikora, a cancer biologist at the University of Colorado, Denver. He uses breast cancer cell lines to work out the effects of 'weak oestrogens', which include certain drugs and industrial chemicals such as bisphenol A. Sikora buys serum that has been treated with charcoal to strip out steroid hormones and other greasy molecules. Then he tests the serum on cells that have or lack oestrogen receptors; if the hormones in the serum have been effectively removed, the proliferation rates should be the same.

Last year, he and others in his laboratory were stalled for about six months when sequential batches of serum failed this initial screen. Differing hormone content "totally flipped" the interpretation of how a cancer drug worked³. Sikora thinks that unrecognized variation in serum might explain why he and a potential collaborator could not get consistent results.

But even when researchers do batch testing, they don't always know what to look out for. Some laboratories simply buy the serum lot in which their cells grow the fastest. Instead, they should tailor screens to the intended study. Researchers also need to report exactly how they screen serum to enable others to reproduce the work, says Sikora.

Some cells and experiments will be more sensitive to the effects of serum than others. The 'transformed' cell lines selected over decades for robust growth tend to vary less than 'diploid' lines or primary cell lines that more closely resemble natural tissue.

Researchers always need to be careful, says Mariella Simon, a cell and developmental biologist at the Children's Hospital of Orange County in California. Ideally, they should have enough serum to last an entire study. And when they do move to a new bottle, they should make sure that no other reagents have changed and that they have enough old serum stockpiled to test whether any strange results can be attributed to the switch. It is easy, for example, to conclude that something is going wrong with a protocol to introduce DNA into cells when, in fact, a new batch of serum has affected division rates. Researchers should also record the information supplied by vendors about serum, including lot numbers, says Simon. "You can't just use your labmate's serum that might have been aliquoted a long time ago and labelled FBS."

BOVINE SERUM'S WIDE RANGE

The bioactive compounds in fetal bovine serum can vary dramatically from lot to lot. Selected components are shown below.

Components	Average (range)
Endotoxin	0.356 ng ml ⁻¹ (0.008–10.0)
Total protein	3.8 g dl ⁻¹ (3.2–7.0)
Alkaline phosphatase	255 mU ml ⁻¹ (111–352)
Lactic dehydrogenase	864 mU ml ⁻¹ (260–1,215)
Cortisol	0.5 µg dl ⁻¹ (<0.1–2.3)
Insulin	10 µU ml ⁻¹ (6–14)
Parathyroid hormone	1,718 pg ml ⁻¹ (85–6,180)
Progesterone	8 ng dl ⁻¹ (<0.3–36)
Testosterone	40 ng dl ⁻¹ (21–99)
Prostaglandin E	5.91 ng ml ⁻¹ (0.5–30.5)
TSH	1.22 ng ml ⁻¹ (<0.2–4.5)
FSH	9.5 ng ml ⁻¹ (<2–33.8)
Growth hormone	39.0 ng ml ⁻¹ (18.7–51.6)
Prolactin	17.6 ng ml ⁻¹ (2.00–49.55)

TSH, thyroid stimulating hormone; FSH, follicle stimulating hormone

Contaminants can confound experiments, too. One of the most insidious is *Mycoplasma*. This tiny bacterium can slip through sterilizing filters and is unfazed by many antibiotics. It depletes cells' nutrients and alters DNA and protein synthesis. An analysis of nearly 10,000 rodent and primate samples found that more than 10% contained RNA sequences unique to *Mycoplasma*⁴. Conventional *Mycoplasma* testing can take several weeks and still miss rare strains, but PCR-based tests are now providing swifter, surer answers, says Yvonne Reid, who leads standards-setting efforts at American Type Culture Collection, a non-profit repository for cell lines in Manassas, Virginia.

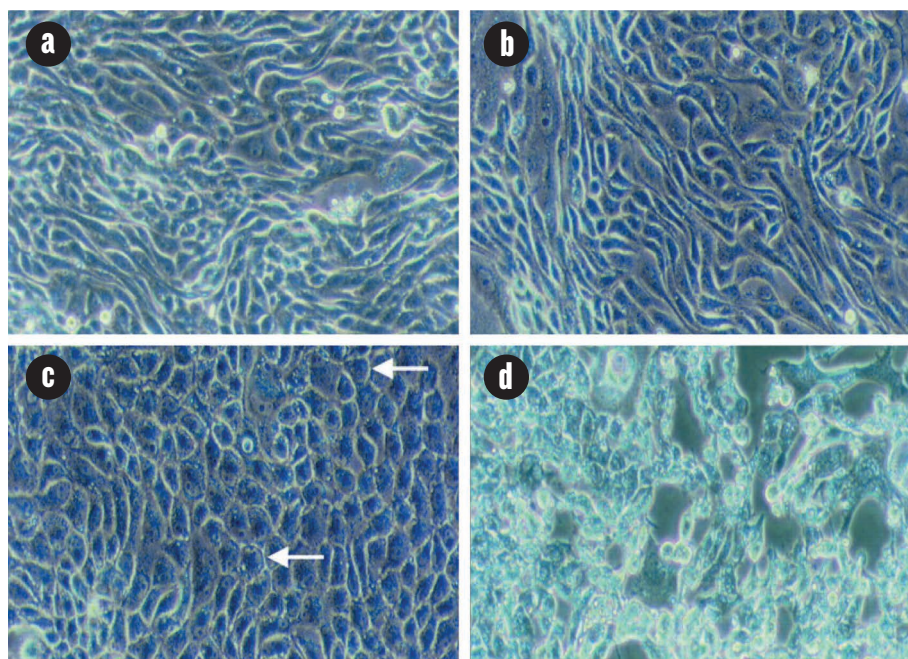
To avoid serum contamination, some researchers are opting for gamma irradiation. Several common contaminants, including *Mycoplasma*, are sensitive to even low levels of radiation. But this requires a balancing act: radiation also damages growth proteins and bioactive molecules that help cells thrive. Many vendors offer gamma-irradiated serum, and the International Serum Industry Association has set up a working group to elucidate its effects⁵. Cell-culture consultant Raymond Nims of RMC Pharmaceutical Solutions in Longmont, Colorado, advises anyone who plans to work with gamma-irradiated serum to first test that cells perform as expected, and to remember that even contaminant-free serum cannot prevent infection by other sources.

ERRORS COME FROM EVERYWHERE

The cells' physical environment is a profound influence. Researchers at the Wyss Institute in Boston, Massachusetts, found that mechanical peristalsis-like deformations and fluid flow changes alone could, without any alterations to the growth media, induce functional villi from cells that otherwise grow flat⁶.

Lab dishes of different brands leach different chemicals into cell-culture media, and can confound studies of cell metabolites. Deliberate additives can change cell metabolism in unappreciated ways: antibiotics in particular frequently impair mitochondrial activity. Even a glass door on a lab refrigerator can ruin experiments, because some chemicals in growth media are sensitive to light. Just changing the laboratory plates, and thus the height of media in which cells are sitting, can alter how cells behave. What's more, cells growing in a given culture are not identical, and the subset of cells that thrives the most can quickly dominate a population. That means cells may not revert back to former behaviour if a researcher decides to restore previous experimental conditions.

In all these experiments, the cells themselves are the most important variable. There is no quick, simple way to know that cells are fit for purpose, says John Masters, a cancer researcher at University College London, and author of cell-culture reference books. "Get to know your cells," he urges. "The best assay you



Epithelial cells growing in regular culture medium (a, at 6 hours; b, at 48 hours) become rounded in the presence of cholesterol (c, shown by arrows) and shrivel and die when cholestane is added (d)⁷.

have for knowing how happy the cells are is looking at them."

Leland Foster, a cell-culture consultant in Salt Lake City, Utah, and former chief executive of HyClone Laboratories, the cell-culture reagents company now owned by GE Healthcare Life Sciences, thinks that trainees cannot, generally, be expected to take the care required. "One way that labs could get away from variability is having some expertise that is resident in that laboratory," he says. Growing cells is, he says, best left to "an expert cell culturist" who can tell when cells are "smiling or frowning", and who will ask serum and other vendors tough questions about the products they buy for their cells.

Cells react differently when they are growing rapidly or persisting in a stationary phase. If they are 'overpassed' (that is, kept in culture too long), other changes can occur and affect reproducibility. Even when the genetic identity of a cell line has been authenticated — as is now broadly recommended — other crucial attributes, such as the growth state, number of doublings and checks for contamination, too often remain undocumented. "Authentication means more than identity," Reid says. Researchers hoping to reproduce experiments should not have to "act like a detective" to work out what state the cells were in when building on a reported study.

Given these unknowns, researchers should take a week or so to optimize their cells' growth and plot a growth profile before launching experiments, says Reid, who is coordinating an open-access series about best practices in cell culture. A growth profile can inform researchers when to harvest cells, when to do assays and when to go back to a distribution bank for a

fresh batch of cells. It can also warn scientists if they are overlooking important variables. Most of all, researchers must be alert and creative to make sure the cells they are using are consistent across a study, Neve says. "There is no single set of experiments that works for everyone."

Concrete data, like good microscope images or expression data, can help researchers recognize when the cells used in their experiment have changed, says Anne Plant, a division chief at the US National Institute of Standards and Technology in Gaithersburg, Maryland, who hopes to find quantitative ways of making cell-culture experiments comparable across laboratories. "One of the hardest things to assess is what constitutes a healthy cell," she says.

Even harder can be the consequences for researchers who neglect to think of cells as "live beings that need to be looked after and cared for," says Masters. "Someone's PhD goes down the pan, or a grant is lost, or years of work are wasted because they are not doing fairly simple quality control." ■

Monya Baker writes and edits for *Nature* in San Francisco, California.

1. Khodabukus, A. & Baar, K. J. *Cell. Biochem.* **115**, 2198–2207 (2014).
2. Hughes, P., Marshall, D., Reid, Y., Parkes, H. & Gelber, C. *Biotechniques* **43**, 575–586 (2007).
3. Sikora, M. J., Johnson, M. D., Lee, A. V. & Oesterreich, S. *Endocrinology* <http://dx.doi.org/10.1210/en.2016-1297> (2016).
4. Olarerin-George, A. O. & Hogenesch, J. B. *Nucleic Acids Res.* **43**, 2535–2542 (2015).
5. Versteegen, R., Plavsic, M., Nims, R., Klostermann, R. & Hemmerich, K. *BioProcess. J.* **15**, 5–11 (2016).
6. Kim, H. J. & Ingber, D. E. *Integr. Biol.* **5**, 1130–1140 (2013).
7. Seo, D. W., Choi, H.-S., Lee, S. P. & Kuver, R. *Am. J. Physiol. Gastrointest. Liver Physiol.* **287**, G1247–G1256 (2004).

CAREERS

POISON PIPES What drives an activist engineer? **p.439**

TRANSFERABLE SKILLS What are scientists good at? go.nature.com/2cabnfp

NATUREJOBS For the latest career listings and advice www.naturejobs.com

MONTY RAKUSEN/SPL



Industrial experience can open doors to whole new career options.

INDUSTRY

Open for business

Postdoc positions in industry can teach people skills that they would not learn in academia.

BY CHRIS WOOLSTON

For better or worse, a postdoctoral position (or two or three) has become a near-mandatory stop en route to a permanent research career. As scientists search for postdoc opportunities, many have had to rethink the template for what constitutes a suitable position. The usual posts at universities or government-run research institutes still attract plenty of applicants, but many researchers are opting to continue their training at a different kind of institution — one with a chief executive instead of a dean.

A postdoc at a for-profit company can open doors to all sorts of science careers.

But just like at universities and institutes, industry postdocs can bog people down in go-nowhere positions — in fact, the industrial realm holds special pitfalls for those who don't carefully check the job requirements and limitations. Before applying for an industrial postdoc, researchers should make sure they will emerge with the skills, publication history and network that they'll need to take their next career step.

Even for those with a deep interest in pharmaceuticals and biotechnology, an industrial postdoc can be far off the radar. That was the case for Nuria Sancho Oltra. After finishing a PhD in organic and biomolecular chemistry at the University of Groningen in the

Netherlands, she took a postdoc position that included two years working on drug development at the University of Pennsylvania in Philadelphia and more than a year at the Swiss Federal Institute of Technology in Lausanne. She hadn't thought of doing a postdoc in industry, but quickly realized that the academic route wasn't for her. "I wasn't curing a disease or doing anything that would improve health care in the short term," she says. "I was publishing papers and that was it."

As she wrapped up the postdoc, she decided she wanted to become a full-time scientist at a drug firm. "I started applying for jobs, but I realized it would not be easy because I lacked industry experience." So, instead ►

POSTDOC APPLICATIONS

How to get your CV noticed

Only stand-out applicants have a real shot at a postdoc position at a top research company. So how to stand out? Sarah Hymowitz, who sifts through hundreds of applications for every postdoc opening in the department of chemistry and structural biology at Genentech in South San Francisco, California, has some suggestions. She doesn't have much time to scan the CVs — and some warrant little more than a glance — so she looks for specific things.

● **Defined purpose.** Hymowitz looks for people who have a specific scientific reason for seeking a position at the company. “A lot of second-tier applicants simply want to work at Genentech,” she says.

● **Ability to finish.** “I’m looking for people who have a history of finishing projects,” she says. She is therefore less than impressed by a list of ‘submitted’ papers on a CV. “An actual paper in *Nature Structural Biology* is better than a hypothetical paper in *Nature*,” she says.

● **Team spirit.** Hymowitz looks for scientists who embrace teamwork, a crucial part of industrial work. “I like seeing middle-author papers,” she says. “It shows you can collaborate.”

● **Science, not business.** Don’t waste precious CV space detailing your business knowledge of biotech or pharma. When hiring postdocs, Hymowitz is first and foremost looking for scientists, not business partners. “I don’t care what they know about industry,” she says.

● **Clear markers.** Hymowitz doesn’t have time to read every CV from top to bottom, so the key info needs to jump off the page. She recommends a couple of bullet points that highlight scientific skills and accomplishments, complete with keywords.

● **Testimonials.** A word of support from someone familiar with your work can go a long way. “If your PI sends me an e-mail or gives me a call, I’ll take a closer look at the application,” Hymowitz says. **C.W.**

► of sending out more futile applications for permanent work, she started a two-year postdoc at the Swiss pharmaceutical company Roche. Four months in, she’s already picked up a lot of industry knowledge about the ideas, experiments and tinkering needed to turn an interesting compound into an actual drug. “I have a more-global vision of what it takes to develop products,” she says. “You interact with so many people. You feel like you’re part of the team.”

INITIAL STEPS

Looking back, she’s happy with her path: she says she wouldn’t have been able to get the practical, health-care-focused post she has now without the training from her academic position. Still, she encourages other scientists with an interest in pharma or biotech to streamline the process and consider an industrial postdoc as their first option. “If you are finishing up a PhD, you are perfectly capable of doing a postdoc in industry,” she says. “It’s not much different from research in academia.”

And the truth is, most companies are reluctant to hire permanent staff who don’t have any industrial experience, says Barbara Preston, a former pharmacologist and the co-founder of PharmaScouts, a science recruitment firm in La Jolla, California. “Companies tell me that it takes a year for people to psychologically make the transition from academia to industry,” she says. Researchers who have an industrial postdoc on their CV are much more attractive to company hiring committees.

Developmental biologist Daniel Lafkas effectively dismissed the idea of an industrial postdoc as he finished up his PhD at the National and Kapodistrian University of Athens. “I thought that if I wanted to do basic research, my only option was academia,” he says. “I wasn’t aware of the level of science conducted at biotech companies.” His plans — and his preconceptions that industry wouldn’t be the right arena for fundamental research — changed after he spoke to cancer researcher Chris Siebel while at a conference. Siebel, a leading figure in oncology at Genentech in South San Francisco, California, shared his commitment to basic research, so Lafkas quickly reevaluated his concept of an acceptable postdoc position. “His standing in the field was a very important factor for me even considering a postdoc in industry,” Lafkas says.

Like many scientists contemplating a stint in industry, Lafkas worried that the corporate culture of secrecy would cut him off from the research community. “You need connections,” he says. “If you can’t go outside of the company to talk about your work, that can be a deficit.” Publications were another key issue, he says. “I knew I had to go into a lab that would allow me to publish well.”

Those concerns are valid, Preston says, who adds that many postdocs in industry are held back by the company culture. “Postdocs want to be able to publish,” she says. “But in industry, a lot of times you can’t.” Some companies are reluctant to publicize their research, and some simply don’t have the funds to support

the sort of side projects that can lead to papers, she says.

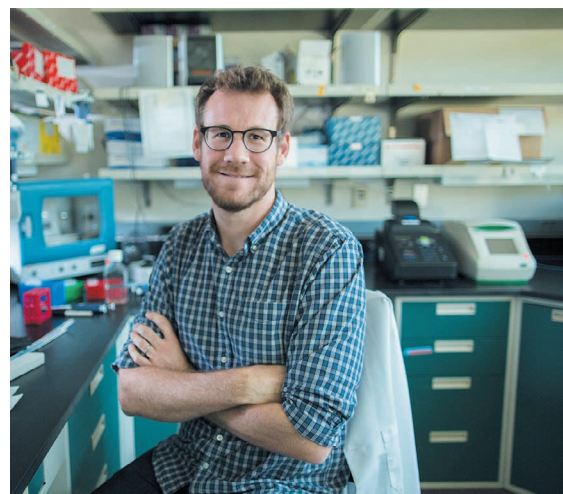
Genentech expects its postdocs to publish, however, and after getting that assurance, Lafkas took a postdoc position in Siebel’s lab in 2013. The move paid off. In 2015, Lafkas was the lead author of a *Nature* paper showing that Notch signalling pathways can determine the development of adult lung cells (D. Lafkas *et al. Nature* **528**, 127–131; 2015). With a paper in a prestigious journal under his belt, he felt he had many options when his postdoc ended in 2016. “Going back to academia was still a possibility,” he says. But he ended up accepting a full-time position in Genentech’s department of immunology discovery, where he’ll join the search for new drug targets. “I wanted to find a lab that would get me out of bed in the morning,” he says. “As long as I’m doing work that I find exciting, I don’t see a need for a change.”

RESEARCH FIRST

New graduates considering their postdoc options may worry if they do their training in industry, they’ll never be able to get back into academia. Although it’s true that most researchers who take industrial postdocs end up staying in industry, that’s far from the only possible outcome, says Leslie Pond, head of the postdoc programme at the Novartis Institutes for BioMedical Research in Cambridge, Massachusetts. “The way our programme is structured, it’s possible to build a path toward an academic career,” she says. “The emphasis is on basic research, and they have the opportunity to establish a strong publishing record.” Novartis also understands that postdocs need to be able to discuss their projects with other scientists, she adds. “Because it’s a temporary position, they need to be able to be specific about the work they’ve done in their future job interviews.”

Pond says that about 5% of Novartis postdocs go straight to full-time positions in

GENENTECH



Daniel Lafkas was initially sceptical that industrial postdocs could incorporate basic research.

academia. Another 8% go on to do a second postdoc, many in academia. Recent alumni of the Novartis postdoctoral programme include Sereina Riniker, a chemist now at the Swiss Federal Institute of Technology in Zurich, and Andreas Bender, a principal investigator working on molecular informatics at the University of Cambridge, UK.

Preston says that scientists who complete a sound industrial postdoc should be well prepared for a career in academia. The main strike against them, she says, is that they won't gain much experience in writing grant applications, which is important for academic survival. Joe Arron, director of immunology at Genentech, agrees that people who do industrial postdocs usually have that important gap in their skill set. "They're coming out of their postdoc without a foot in the money bucket," he says. "Typical academic postdocs are going to be more involved in the grant process." It's always possible to learn how to write grant applications through seminars, workshops or online courses, however, and Genentech offers its employees special grant-writing programmes.

It's understandable that industrial postdocs tend not to return to academia, Preston says. Certain personalities are simply better suited for industry, and those who thrive there are likely to want to stay. "In industry, you have to be team-oriented and cooperative," she says. "People in academia are more independent."

Cooperative or not, it takes a competitive edge to get in the door at a top research company. Arron says that he gets hundreds of applications whenever there's a postdoc opening in his lab. "We're looking for really great scientists with a lot of potential," he says (see 'How to get your CV noticed'). "Beyond that, it's open-ended."

In his experience, many of the top scientists didn't have a clear preference for academia or industry when considering their postdoc options. Instead, they were looking for the right mentor with the right project, no matter where it might be. "If you're a talented scientist, you want to go to an elite institution in your area," he says. "We're competing with top academic and medical centres for postdocs."

In the end, Arron says, industrial postdoctoral positions can be just as valuable and productive as academic postdocs, and vice versa. "Good science," he says, "is good science." ■

Chris Woolston is a freelance writer in Billings, Montana.

TURNING POINT

Activist engineer

Last year, civil engineer Marc Edwards spent at least US\$150,000 of his own money to prove that tap water in Flint, Michigan, was contaminated with lead. Over the past decade, Edwards has been documenting and exposing lead contamination in the Washington DC water supply and fighting to hold government officials accountable. Edwards explains how this work equipped him for the Flint case, which garnered international attention and shone a spotlight on similar concerns nationwide.

A mother's plea for help got you involved in the Flint crisis. Is it similar to the DC case?

In Flint, up to 12,000 children have been exposed to high lead levels. The DC-area case was much worse than Flint, in terms of harm done and number of children affected. Unfortunately, there was betrayal by government officials in both cases.

How did the DC case prepare you for Flint?

As a civil and environmental engineer at Virginia Polytechnic Institute and State University in Blacksburg, I researched corrosion in homes. In 2003, I started sampling water in DC homes and found outrageously high levels of lead. Ultimately, we discovered that the public had been misled by local and federal agencies. I've had to disprove falsified government reports, which my earlier work had not prepared me for. But without that experience, I would not have been able to help people in Flint.

How did the events in Flint unfold?

Flint was the exact opposite of DC in every respect. Once we confirmed the contamination and government oversight, we had sample kits going to Flint in less than a week. We knew we had to cooperate with anyone who wanted the truth about the lead, and fight anyone who tried to obfuscate matters. There is a line between science and activism, and it's one you cross only as a last resort. It's either that or, in this case, letting kids be hurt and a city destroyed. We used Freedom of Information Act (FOIA) requests — which invoke a federal law to access information from the government — to get the data about who knew what was happening with the contamination and when.

Your findings contradicted official reports.

Were you concerned about credibility?

Only the paranoid could possibly survive something like this. If you make one mistake, you will never, ever recover. It makes you very careful not to say anything you are not prepared to back up 100%.



How have your efforts affected your workload?

I worked on the DC case for 30 hours a week as a volunteer, for 10 years. But I worked 70 hours a week to make money and produce papers, the things that count towards academic-career success. There's no way you'd put on your CV that you made FOIA requests and attempted to get falsified reports retracted.

How did you fund the Flint work?

I knew the day would come when another community would need help, so I donated my fees from consulting and other work into a fund in the department. It was put into a discretionary account. We did, eventually, get \$33,000 from the US National Science Foundation, which gave us credibility.

Are you getting calls from people in other cities about more contamination concerns?

I get 20–30 communications every single day. I work 65 hours a week on Flint, so I don't have time to check these things out. But in the back of your mind, you say, what if they are valid?

Why do you maintain a website with Flint research updates?

I didn't want to be dependent on the few investigative reporters left to explain the science behind it. Every single major breakthrough came out on our blog first.

Do you have lasting concerns?

There was a time when engineers and scientists were the leaders of their generation. But we have created our own world, set apart from society, where we tell each other we're important. If we cannot get this fixed, we are destined to enter a new dark age. ■

INTERVIEW BY VIRGINIA GEWIN

This interview has been edited for length and clarity.